THESIS

CHARACTERIZING DISTRIBUTIONS AND DRIVERS OF EMERGENT AQUATIC VEGETATION IN MINNESOTA

Submitted by

Jillian LaRoe

Graduate Degree Program in Ecology

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Summer 2020

Master's Committee:

Advisor: Jody C. Vogeler

Wade Tinkham
Paul Evangelista

ABSTRACT

CHARACTERIZING DISTRIBUTIONS AND DRIVERS OF EMERGENT AQUATIC VEGETATION IN MINNESOTA

The emergent aquatic vegetation (EAV) communities across the lakes of Minnesota serve critical functions within ecosystems by providing habitat and forage for native waterfowl and fish species, moderating water chemistry, and serving as a cultural and economic resource. Communities of EAV are changing dramatically in response to alterations in hydrologic flow regimes, nutrient availability, biological homogenization, and near-shore development. To address the conservation of these communities at a spatial scale relevant for landscape management, the changes need to be evaluated at local and regional scales. Previous efforts to map and monitor EAV have utilized field surveys, aerial imagery, multispectral imagery, and synthetic aperture radar (SAR). However, it is difficult to apply the findings of previous studies to broader spatial scales because they lack field surveys, clear or repeatable methodologies, rigorous validation, and/or applying methods to broad spatial extents, all of which are all necessary for providing direct implications for landscape level management. The first chapter of this thesis aimed to overcome these challenges and create statewide maps of EAV in Minnesota at a spatial resolution relevant to landscape management at both broad and local scales. We paired detailed field surveys of EAV communities with Sentinel-1 SAR and Sentinel-2 Multispectral Imager to create annual maps of EAV across the lakes of Minnesota at a 10 m spatial resolution in 2017 and 2018. We created two random forest models, a species model predicting general classes of EAV and a water model identifying open water regions across hydrologic features in Minnesota. We validated both classification models using withheld field sample locations to measure overall accuracy as well as individual class user's and producer's accuracies. The species and water map predictions were combined into a final map representing water and EAV classes each year. We also evaluated each map by the area-based percentage of overlap between model predictions and field surveys

which ranged from 54.5 to 90.1% agreement. The 2017 map was further evaluated using an area-based weighted probability with an overall accuracy of 89.9% (±0.7%). The methods and promising results highlighted by this study set the stage for subsequent analyses at broader spatial scales to quantify temporal shifts or trends in EAV communities. The combination of these diverse and detailed datasets provides methods for generating annual maps of EAV distribution across Minnesota, and ultimately provide a tool to support landscape-scale conservation efforts of EAV communities in Minnesota.

The second chapter investigated the influence of systemic drivers related to the decline of northern wild rice (*Zizania palustris L.*) over the last century. Wild rice is an environmental indicator species that is sensitive to hydrologic changes and disturbances and serves an essential role in ecological, cultural, and economic systems in Minnesota. Due to the previous lack of comprehensive information regarding its extent and distribution, previous efforts to study its decline have been limited to small regions or small samples of lakes across the state. We utilized 2018 presence maps of wild rice from the first chapter and summarized wild rice cover across 366 lakes. Then, we employed a suite of spatial, hydrological, ecological, and environmental variables summarized at a variety of spatial scales within a three-step modeling framework to select the most significant drivers of wild rice cover, explore interactions between drivers, and account for inherent spatial autocorrelation in the datasets. A final spatial lag model revealed that dispersal and population connectivity had the strongest relationships with wild rice cover on each lake. While further exploration may better quantify this relationship, land managers should consider the degree of connectivity between wild rice lakes and their spatial configuration on the landscape during conservation planning to maximize population resilience. Our results suggest that it may be more suitable to approach populations as connected habitat regions, in contrast to the more widely accepted notion that wild rice lakes are self-contained or independent populations.

ACKNOWLEDGEMENTS

The Minnesota Department of Natural Resources (MN DNR) was an essential collaborator in the research completed here, and I thank them for their input and strong interests, especially Paul Radomski and Josh Knopik for their higher level of involvement with the research process and providing data access. MN DNR's excellent records and intensive field data surveys collected with great precision and high accuracy empowers this research to be of higher caliber, with greater opportunistic potential since models will only be as good as the data used to train them.

I would also like to thank my advisor Jody for all of her help and support through this process, and encouraging me to pursue research that I am most passionate about. Her guidance shaped this research and substantially improved the rigor of work produced. I would like to recognize the contributions from my committee members Wade and Paul for their insight and guidance at various stages of this work. I would also like to thank Steve Filippelli, Eric Jensen, Patrick Fekety, Tim Mayer, Tony Vorster, Nick Young, Kristen Dennis, and many other folks that I have worked with through NASA DEVELOP and NREL over the years. Their individual contributions and support at many different stages has made this work possible and significantly promoted my growth and development as a research scientist, ecologist, and spatial data analyst.

I could not have done this without the help of my partner Neal Swayze who supported me at every step through graduate school while completing his own thesis as well. Each year presented immense obstacles in life and learning, and I am so grateful that we did this together, supporting one another.

TABLE OF CONTENTS

CHAPTER 1: MAPPING EMERGENT AQUATIC VEGETATION IN MINNESOTA

Introduction

Background

Emergent aquatic vegetation (EAV) communities are prevalent within shallow water and banks across the expansive lakes in Minnesota and they represent a group of functionally critical species within aquatic ecosystems (Radomski & Goeman, 2001; Wilcox & Meeker, 1992). EAV are essential to the diets and habitats of native fish, waterfowl, and other species, and also have direct influences on water chemistry (Joniak, Kuczyńska-Kippen, & Nagengast, 2007; Radomski & Goeman, 2001). Within the last century, habitat loss, environmental degradation, invasive species, and changing climate have notably contributed to shifts in species richness and community composition of EAV on many lakes (Alahuhta, 2014; Pillsbury & McGuire, 2009; Hansel-Welch et al., 2003). The spatial patterns in species richness and composition of aquatic vegetation communities have been explained by local and regional scale characteristics, most notably water quality and climatic variables; specifically, alkalinity, concentration of phosphorus, and temperature, varying with latitude throughout Minnesota (Alahuhta, 2014; Moyle, 1956). The large overarching trends in EAV populations and communities have not been explained solely by local factors, which highlights the critical need to consider conservation and management at appropriate spatial scales that are broad enough to capture potential regional abiotic gradient drivers (Alahuhta, 2014). Recent studies in ecology and conservation have emphasized the necessity of evaluating changes in populations and communities at multiple scales (i.e. evaluating a collection of ecosystems) due to their interconnectedness and spatial dependence on surrounding features (Forman & Gutzwiller, 2002). Conservation and monitoring efforts traditionally utilize field surveys as a primary information source; however, current approaches are constrained by the lack of comprehensive information available about the distribution and temporal trends of EAV communities at a spatial extent relevant to these communities and species in particular (Reinke & Jones, 2006; Forman & Gutzwiller, 2002). Outlining strategic and refined methods for comprehensively mapping EAV distributions on an annual basis

is a necessary first step toward ensuring the preservation of EAV communities across the lakes of Minnesota by improving the information available for proactive land management decisions and conservation efforts.

Species Composition

Most EAV species are perennial plants growing in shallow water or along the banks of water bodies. Common species in the region include cattails (*Typha* spp), water lilies (*Nymphaeaceae* spp), bulrush (*Schoenoplectus* spp), wild rice (*Zizania palustris L.*), watershield (*Brasenia* spp), and sedges (*Carex* spp) (Muthukrishnan & Larkin, 2020). Their abundance and community composition are related to characteristics such as hydrologic flows, lake morphology, sediment composition, nutrient availability, water clarity, and upstream land cover, many of which are affected by urbanization and land use changes (Kissoon et al., 2013; Joniak, Kuczyńska-Kippen, & Nagengast, 2007).

Cattails are one of the most common perennial species found within marshes and along shallow banks of water bodies. There are several cattail species like broad-leaved cattail, narrow-leaved cattail, and a hybrid between these two species found throughout Minnesota (Larkin et al., 2011; Dubbe, Garver, & Pratt, 1988). They are capable of dominating over sedges and other wetland grass under the appropriate conditions, and encroachment has been attributed to changes in hydrologic flows and water levels (Wilcox et al., 2008). Cattail growth typically begins between April and May and tapers out by October, although this varies by region (Dubbe, Garver, & Pratt, 1988; Apfelbaum, 1985). Cattails experience more rapid growth during their second year, after they have established, and the majority of this happens between July and September (Dubbe, Garver, & Pratt, 1988). Sedges and bulrush are also found growing within shallow waters, and the primary species include hard-stem and soft-stem bulrush which are perennials growing up to 3 m in height (Price, 2012).

Northern wild rice (*Zizania palustris L.*), or wild rice, is found growing along the edges of water bodies concentrated in the northern half of Minnesota at depths less than 1.5 m (Price, 2012). It is considered an indicator species due to its highly sensitive and responsive nature to changes in water quality, the environment, and other disturbances, making wild rice key for identifying environmental changes (Biesober, 2019). It has been disappearing from its native range during the last century (Pillsbury & McGuire, 2009), and

is a top priority species within conservation due to its keystone role in wetland ecosystems. The disappearance of historical stands of wild rice from individual lakes has been associated with major shifts in EAV community composition (Pillsbury & McGuire, 2009). As an annual rooted macrophyte, wild rice begins growing from submerged seeds early in the season and transitions to the floating leaf stage where it is visible on the water's surface between mid May and mid June. It then emerges from the water in tall vertical and leafy structures that grow rapidly between mid June and July and continues to grow above the water. When wild rice reaches its peak between August or early September, the seeds disperse via wind or water and wild rice stands senesce until the following year (Price, 2012). Water lilies are another family of rooted macrophytes commonly found growing alongside wild rice at similar water depths (Myrbo et al., 2017(a); Pillsbury & McGuire, 2009). Found floating on the surface of many lakes across Minnesota, water lilies are represented by several native as well as invasive species.

Threats, Drivers, & Trends

Long term monitoring in Minnesota has identified significant reductions in EAV cover through the last 30 years across many lakes (Radomski, 2006). These losses have been attributed to a number of factors including anthropogenic land-use change, water quality, water clarity, alterations in hydrologic flow regimes, climate change, and invasive species. Shoreline development has negatively impacted floating leaf and EAV abundance with lower impacts on submersed vegetation (Pillsbury & McGuire, 2009; Radomski & Goeman, 2001). Anthropogenic land-use change through urbanization and agriculture have resulted in habitat loss and impacted water quality (Hansel-Welch et al., 2003); these are key factors that have caused increased sediment loads which reduce water clarity, and increased runoff with high concentrations of nitrogen and phosphorus, where the negative impacts accumulate downstream (Blann et al., 2009). Human impacts have been attributed as a causal factor of elevated pore water sulfide concentrations which is toxic to wild rice and serves as one imperative component responsible for unprecedented changes in community dynamics (Myrbo et al., 2017(a); Pillsbury & McGuire, 2009). Alterations in hydrologic regimes have resulted in significant changes to annual flow rates, channelization of streams, and countless consequences for riparian and aquatic biodiversity (Blann et al., 2009). Climate change has impacted temperature and precipitation regimes, altering suitable habitat for

species and creating conditions that favor the influx of invasive species (Rahel & Olden, 2008). Biotic homogenization is noted to be correlated with increased abundance of invasive species presence in studies conducted across more than 1,000 shallow lakes in Minnesota, and the impacts have repercussions for EAV communities (Muthukrishnan & Larkin, 2020). These changes may dramatically alter and limit the future abundance of EAV, increasing the need for information regarding current species and community distributions at relevant spatial scales to inform proactive management and preserve the essential roles EAV play in ecosystem function.

Previous Monitoring Efforts

The earliest efforts to monitor EAV and wetlands primarily utilized field surveys and aerial imagery across multiple decades, yet these methods have proven to be challenging as well as time and resource intensive to collect (Wilcox et al., 2008; Rundquist, Narumalani, & Narayanan, 2001). Studies have demonstrated that it is infeasible to comprehensively collect field surveys across large regions and dense vegetation, necessitating the incorporation of aerial imagery (Randomski, 2006). Through advancements in temporal, spatial, and radiometric resolution of satellites, previous literature has investigated methods for classifying aquatic vegetation using multispectral imagery (Villa et al., 2018; Price, 2012; Sawaya et al., 2003; Rundquist, Narumalani, & Narayanan, 2001). In particular, the Landsat archive represents a long running temporal record of spectral imagery at resolutions that may be relevant for characterizing vegetation patterns. However, Landsat imagery alone has proven limited in its effectiveness for monitoring EAV, likely due to difficulties penetrating cloud cover, capturing below canopy vegetation, and differentiating between vegetation species at fine scales (Mansaray et al., 2017). Therefore, previous studies have explored supplementary spectral imagery and data sources to improve mapping efforts for EAV communities of conservation interest (Villa et al., 2018; Mansaray et al., 2017; Price, 2012).

Synthetic aperture radar (SAR) is an active remote sensing technique emitting microwave signals that reflect off the landscape surface and return to the sensor known as backscatter (Dabboor & Brisco, 2019). The backscatter contains information about structure of land surface features, stored using single and dual polarizations that capture vertical and horizontal surface structure (Silva et al., 2007). SAR is an exceptionally

powerful tool for wetland classification and monitoring emergent vegetation because it captures the height, density, and spatial patterns of wetland and EAV canopies (Dabboor & Brisco, 2019; Brisco et al., 2013; Silva et al., 2007). Studies have incorporated SAR values (or alternatively, applied thresholds to SAR values) to isolate key characteristics of EAV, wetlands, and water and paired this information with multispectral satellite imagery to achieve substantial improvements in classification accuracy (Mansaray et al., 2017; Brisco et al., 2013; Bourgeau-Chavez et al., 2009). Dennis & LaRoe et al. (*in review*) combined Landat 8 OLI imagery with SAR at two time periods in the growing season to predict wild rice distribution within lakes residing in a single Landsat scene with replicable methods. They found that the range of the Vertical-Vertical (VV) polarization across the growing season provided the greatest predictive power, which is also suggested by Gallant et al. (2014). Dennis & LaRoe et al. (*in review*) also suggested that the Vertical-Horizontal (VH) polarization may have improved misclassification between species. While many studies have demonstrated the efficacy of utilizing multispectral imagery and SAR for classifying wetlands or EAV, current classification detail, spatial resolutions, and study extents may be limited in their utility for regional management planning. Many of the previous studies have lacked at least one key element to create EAV maps relevant to land management including field survey data, rigorous validation, clear methods for replicability, applicability through time, or testing methods at a regional scale.

Objectives

The objective of this study was to improve the accuracy of mapping EAV by expanding on remote sensing methods tested by Dennis & LaRoe et al. (*in review*); their study focused on mapping wild rice at a 30 m resolution using Landsat imagery and synthetic aperture radar at two key intra-annual time periods at the extent of a single Landsat scene. Within this study, we incorporate multiple EAV classes for mapping species at a statewide extent using remotely sensed imagery at finer spatial resolutions (10 m), through the delineation of four phenologically significant time periods for greater distinction between classes. We pair imagery captured by Sentinel-1 C-band SAR and Sentinel-2 Multispectral Imager (MSI) with detailed field surveys to generate random forest models to predict five EAV classes as well as open water regions across Minnesota in 2017 and 2018. We highlight the importance of using multiple phenologically significant time periods as well

as the benefits of employing high capacity platforms for analyzing satellite imagery across broad spatial extents across multiple years. The methods outlined in this study provide a robust potential framework for creating annual statewide maps of EAV.

Methods

Study Area

Minnesota is located in the northern portion of the central US, spans over 225,000 km² and contains over 11,000 lakes, thousands of streams and rivers, with surface water representing 8% of all land cover. It is home to the headwaters of the Mississippi River, which provides a variety of ecosystem services to people within 10 states in the US. Natural lakes were glacially formed, and the paleoecological history of the region has molded currently observed trends in water quality and ecosystem processes, such as spatial patterns of pore water sulfide concentrations (Myrbo et al., 2017(a)). Minnesota's diverse landscape covers a number of ecological, geological, and physical gradients from lakes, peatlands, and wetlands, to densely populated cities. Annual precipitation between 1981 and 2019 across the state was 69.5 cm and average temperature ranged from -0.5ºC to 10.8ºC (MNDR, n.d.). The focal time period of this study covers 2017 and 2018, and the cumulative precipitation in each year was 69.1 cm and 72.2 cm respectively (MNDR, n.d.). Mean annual temperature in 2017 ranged from 0.3ºC to 12.4ºC, and 2018 mean annual temperature ranged from -0.8ºC to 10.4ºC (MNDR, n.d.). This study was constrained to the hydrologic features across Minnesota (Figure 1.1).

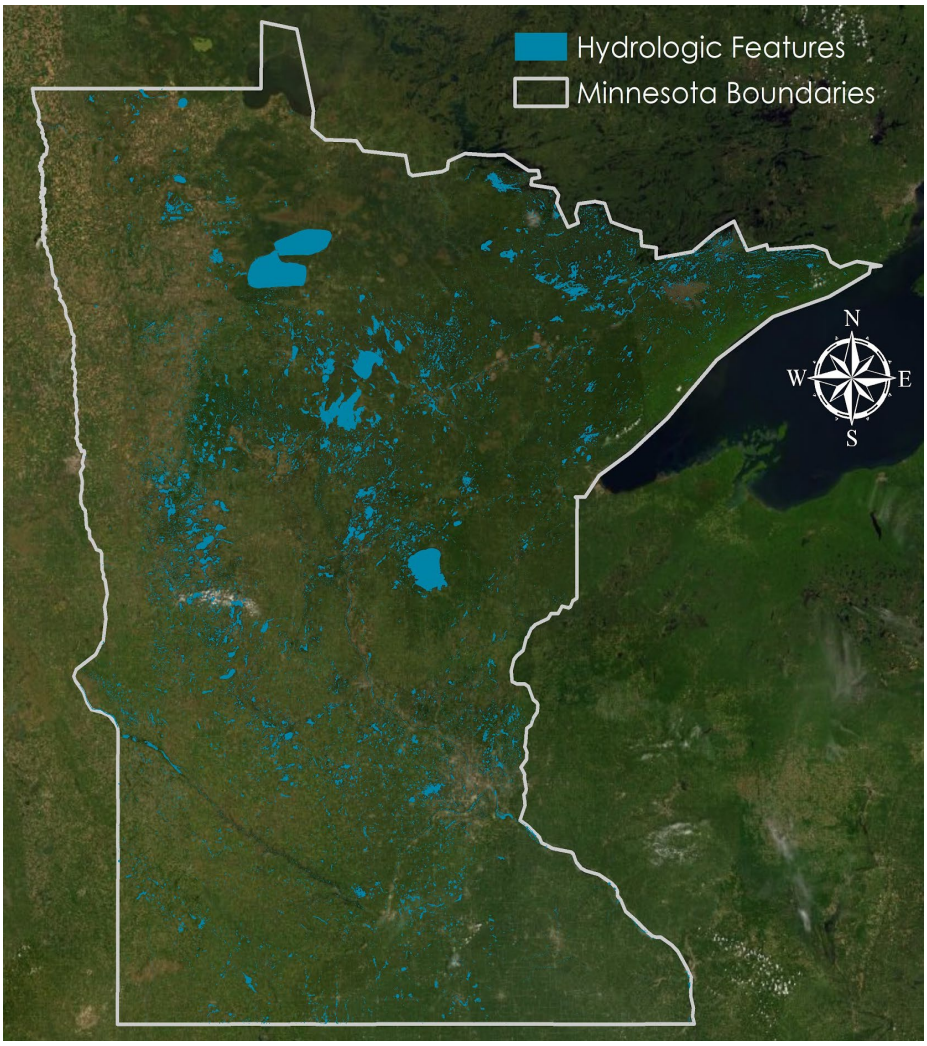


Figure 1.1. Hydrologic features delineating the study area across Minnesota.

Field Survey Data

Detailed annual aquatic vegetation field survey data was collected on a subset of lakes by the Minnesota Department of Natural Resources (MN DNR). These surveys included six categories of taxa with six additional categories of mixed EAV stands classified by the dominant species (Radomski et al., 2011). The survey polygons also contain information regarding any secondary taxon and additional taxa present in a stand of EAV. The primary categories surveyed included cattails, rushes, wild rice, water lilies, submerged vegetation, other emergent vegetation, and other floating vegetation (Table 1.1) and the spatial accuracy of surveys were determined to be between 2 and 3 m (Radomski et al., 2011). Surveys are collected across a

variety of lakes, using different lakes each year to cover a greater total number of lakes in Minnesota through time. These detailed and robust surveys are not possible to complete across exceptionally dense vegetation (Radomski et al., 2011). The sheer number and size of lakes in Minnesota further contributes to the impracticality of surveying all lakes with EAV, and consequently, annual surveys represent a small percentage of the lakes with EAV across Minnesota.

Table 1.1. The survey class definitions for field polygons and the new classification scheme utilized for modeling. The primary taxon is a dominant taxon in a stand of EAV, and a secondary taxon is listed if there is an additional class that covers > 30% of the total stand area. Associated taxa refer to other species present in a stand of EAV, taking up < 30% of the total stand area. Dominant stands (i.e. "Cattails") represent a single dominant taxon, no secondary taxon, and potentially include associated taxa. Heterogeneous stands (i.e. "Cattails and Others") represent a dominant taxon, a secondary taxon, and potentially include associated taxa. Additional details can be found in Radomski et al. (2011).

MN DNR Survey Class	MN DNR EAV Survey Definition	Modeling Class
Cattails	Cattails are the dominant taxon in a more homogeneous stand, and associated taxa may be present.	Cattails
Other Emergent	Other Emergent species are the dominant taxon in a more homogeneous stand, and associated taxa may be present.	Other EAV
Other Floating	Other Floating species are the dominant taxon in a more homogeneous stand, and associated taxa may be present.	Other EAV
Rushes	Rushes are the dominant taxon in a more homogeneous stand, and associated taxa may be present.	Rushes
Submerged	Submerged vegetation is the dominant taxon in a more homogeneous stand, and associated taxa may be present.	omitted
Waterlilies	Water lilies are the dominant taxon in a more homogeneous stand, and associated taxa may be present.	Water lilies
Wild Rice	Wild rice is the dominant taxon in a more homogeneous stand, and associated taxa may be present.	Wild rice
Cattails and Others	Cattails are the dominant taxon in a more heterogeneous stand, a secondary taxon is present, and associated taxa may be present.	omitted
Rushes and Others	Rushes are the dominant taxon in a more heterogeneous stand, a secondary taxon is present, and associated taxa may be present.	omitted
Waterlilies and Others	Water lilies are the dominant taxon in a more heterogeneous stand, a secondary taxon is present, and associated taxa may be present.	omitted
Wild Rice and Others	Wild rice is the dominant taxon in a more heterogeneous stand, a secondary taxon is present, and associated taxa may be present.	omitted

However, these field surveys provide essential data that can be utilized for modeling the distribution of EAV classes across multiple years. While the surveys were not collected with the intention of modeling EAV distributions using remotely sensed data, they can be modified for this purpose (Figure 1.2). Challenges with inflated false positive error rates have been identified within models that incorporated presence points in

which wild rice was not the dominant taxon, or in other terms, using highly heterogeneous EAV stands to represent the presence of a single EAV class (Dennis & LaRoe et al., *in review*). To account for this, we only included the six classes with a single dominant taxon and no secondary taxon present (only associated taxa) to generate presence points, merging other floating and other emergent into a single "other" class (Table 1.1).

We converted survey polygons to 1 m rasters and snapped to Sentinel-2 10 m imagery (Figure 1.2), and used the aggregate sum to count the number of 1 m pixels inside each 10 m pixel to approximate the percent cover of a polygon within a Sentinel pixel. To avoid confusion between mixed classes within the models and retain great enough variance in the density of vegetation cover, we selected pixels with aggregate values ≥ 60 and converted them to training points (Figure 1.2). In other words, a polygon must share 60% overlap with a 10 m Sentinel pixel in order to be considered as a potential training sample. Training samples were selected using a random stratified sample design to balance the number of EAV points included from each class and include nearly equal numbers of training samples from each year (Appendix 1A).

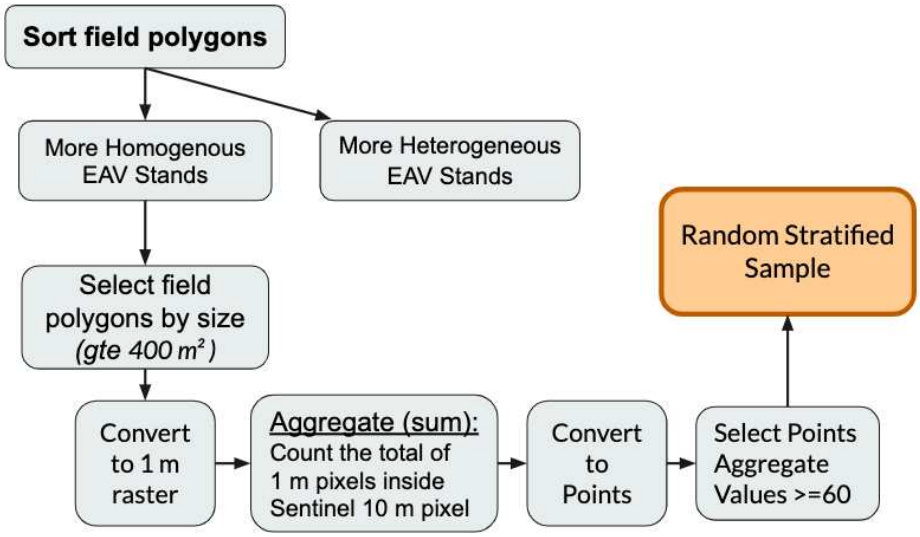


Figure 1.2. Methods for converting field polygons to training and validation samples. The field survey polygons representing more heterogeneous EAV stands were omitted prior to creating and selecting training samples, to refine the training samples to homogeneous EAV stands and provide spectral values more representative of each individual class. The final random stratified sample was split 75/25 for training and validation samples.

The MN DNR field surveys have not traditionally been collected for remote sensing purposes and therefore do not contain information about surrounding or adjacent land cover, and in this case, water was not explicitly included within surveys. To accurately classify EAV and prevent confusion of these classes with

adjacent land cover types, it was critical for us to identify areas containing open water. Open water regions were ocularly sampled using 2017 data from the National Agriculture Imagery Program (NAIP) at a 1 m spatial resolution in Google Earth Engine (Gorelick et al., 2017), which was collected between early July through mid October across the state. To create the ocular samples, we aimed to balance points across the full range spectral range of open water, provide even geographic distribution across the study region, and capture the variation between rivers and lakes of different sizes. Statewide images were displayed in color-infrared to improve visual distinction between water and shoreline vegetation or low-density EAV (10% cover or less). Survey polygons from 2017 were displayed overtop of high-resolution imagery to prevent overlap between vegetation and water training samples. We sampled points using a 20 m grid from median composites of Sentinel-2 imagery overlaid on NAIP to prevent multiple points from being placed within a single pixel. Additionally, the 20 m grid allowed for points to be accurately placed on pixels near lake shorelines containing ≤ 10% cover of vegetation (Figure 1.3). At random, anthropogenic structures like docks and boats were also included in the sampling process to prevent these regions from being misclassified as vegetation. NAIP imagery is typically collected every other year across states, and there was not imagery collected across Minnesota in 2018. For this reason, we did not ocularly sample points from this year. In total, we placed 14,230 points across open water regions in 2017.

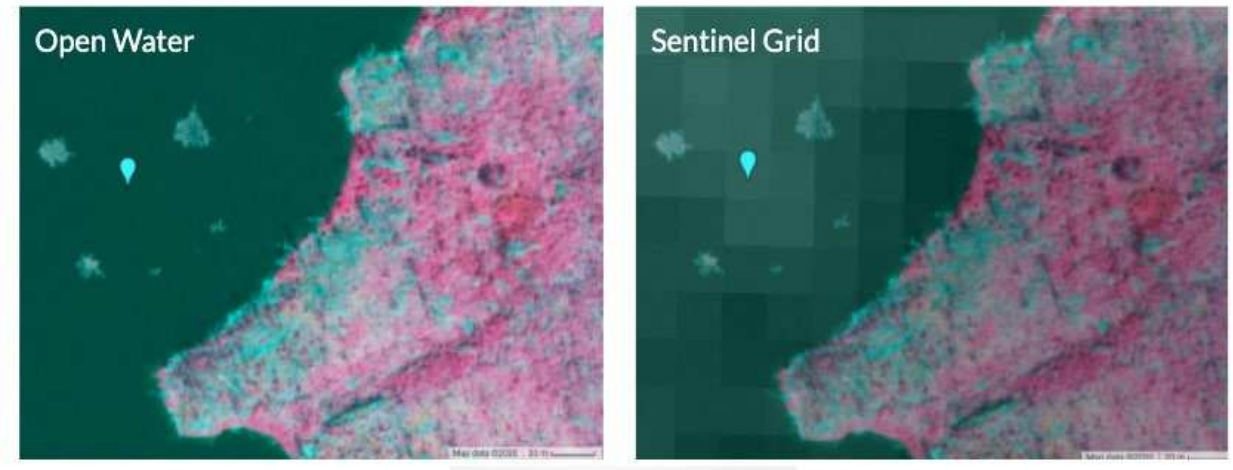


Figure 1.3. Ocular sampling methods utilized to place open water points referencing a Sentinel grid.

Sentinel-1 and Sentinel-2 Preprocessing

We acquired Sentinel-2 MSI imagery using the Google Earth Engine platform (Gorelick et al., 2017) and resampled all spectral bands to a 10 m spatial resolution, closest to the resolution of field surveys (Table 1.2). All Sentinel-2 images collected in 2017 and 2018 at the four time periods of interest (Table 1.3) were cloud-masked and used to create a median image composite at each time period (four composites per year, a total of eight composites). We derived spectral indices related to water and vegetation from median composite images at each time period (Appendix 1B). We selected these four key time periods based on some of the suggestions made by Gallant et al. (2014) as well as a synthesis of relevant literature to isolate key phenological changes, and the specific date ranges for each sensor are shown in Table 1.3. The first time period (T1) captures perennial structures and early season growth between mid-March and the end of May, while the second time period (T2) references the bulk of seasonal growth between June and the end of July. Peak growth and partial senescence is captured in the third time period (T3) between August and late September, and the bulk of senescence occurs between late September and mid November which is the fourth time period (T4).

Table 1.2. Sentinel 2 spectral bands and resolutions retrieved from https://www.satimagingcorp.com/satellite-sensors/other-satellite-sensors/sentinel-2a/

Spectral Predictor	Sensor	Native Spatial Resolution	Central Radiometric Resolution (μm)
B1 Coastal	Sentinel-2	60 meters	0.443
B2 Blue	Sentinel-2	10 meters	0.49
B3 Green	Sentinel-2	10 meters	0.56
B4 Red	Sentinel-2	10 meters	0.665
B5 Red Edge #1	Sentinel-2	20 meters	0.705
B6 Red Edge #2	Sentinel-2	20 meters	0.74
B7 Red Edge #3	Sentinel-2	20 meters	0.783
B8 Near Infrared	Sentinel-2	10 meters	0.842
B8A Red Edge #4	Sentinel-2	20 meters	0.865
B9 Water Vapor	Sentinel-2	60 meters	0.945
B11 Shortwave Infrared 1	Sentinel-2	20 meters	1.61
B12 Shortwave Infrared 2	Sentinel-2	20 meters	2.19

Table 1.3. Phenological time periods and the associated date ranges of satellite imagery selected for each sensor and year. The dates of Sentinel-1 and Sentinel-2 imagery differ slightly; the Sentinel-1 time periods are shorter to isolate key phenological changes since SAR is not affected by cloud cover. Whereas Sentinel-2 is affected by cloud cover, so multispectral image time periods are longer in order to capture enough cloud-free pixels to generate a statewide composite image.

Years	Sensor	Time 1 (T1)	Time 2 (T2)	Time 3 (T3)	Time 4 (T4)
2017 + 2018	Sentinel-1 SAR	4/10 – 5/20	6/10 – 7/15	8/5 – 9/20	9/25 – 11/10
2017 + 2018	Sentinel-2 MSI	3/15 – 5/31	6/1 – 7/31	8/1 – 9/20	9/21 – 11/15

Additionally, we acquired Sentinel-1 C-band SAR at a 10 m resolution in GEE. We generated median SAR composites for 2017 and 2018 at each of the same four time steps as the multispectral imagery (Table 1.3). Appendix 1B provides the calculations used for the range of the median VV and VH values. These calculations were applied between all six possible combinations of the four time steps, for a total of 12 composites depicting the median range of VV and VH values for each year. Additional spectral variables highly correlated with the presence of EAV classes (coastal, blue, and water vapor bands) were also differenced using all possible combinations of the four time periods and included as predictors for model selection. There were more than 120 predictor variables considered within the species model. We extracted all predictor values from both Sentinel-1 and Sentinel-2 at each EAV training data sample for the species model. The same Sentinel-1 and Sentinel-2 predictors utilized for the species model were parsed down to only include the 21 variables solely from the third time period (T3, Table 1.3), which align with the temporal resolution of 2017 NAIP imagery used for ocular sampling. Only the 2017 EAV sampling points utilized in the species model were merged into a single class representing vegetation presence, and they were merged with the ocularly sampled open water points (Figure 1.4). All EAV points and open water points from 2017 were brought into GEE, and the values of T3 predictors from 2017 were extracted for each point.

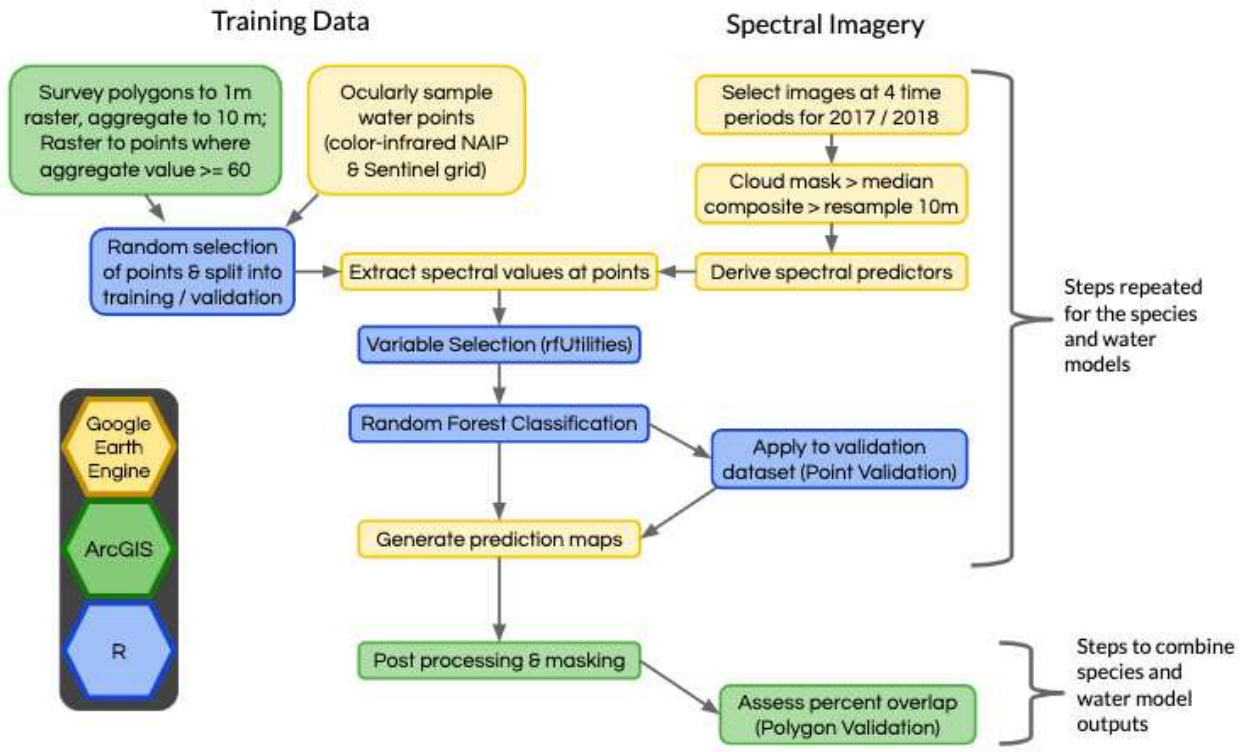


Figure 1.4. Overview of methods workflow for creating and combining the species and water models across Minnesota for 2017 and 2018. Note that ocularly sampled water points were only included in the water model.

Random Forest Models

Species occurrence and open water occurrence datasets were split 75%/25% for training and validation. We employed the rfUtilities package (Version 2.1-2, Evans et al., 2011) in R (Version 3.6.3; R Core Team, 2020) to rank variables by their importance and selected a final set from the top 55 variables. We considered variables based on their degree of explanatory power as well as their ecological significance (relationships with EAV classes) to harmonize human input with the automation algorithmic variable selection (Young et al., 2020). For example, all VV range and VH range variables were included in the species model due to the detailed level of seasonal variance captured by SAR which has proven to be effective for distinguishing emergent aquatic vegetation (Dennis & LaRoe et al., *in review*; Gallant et al., 2014; Brisco et al., 2013). We removed all variables that had a Pearson's correlation coefficient > |0.75|, as well as 60 m resolution variables beyond the top six to concentrate the bulk of the model's explanatory power within variability distinguished at a finer spatial resolution (≤ 20 m). To account for interannual variability in the region, we included data from 2017 and 2018 in the species dataset to train a random forest model (package

version 4.6-12) in R (Breiman, 2001). These methods extended to the water model which only included training samples from 2017, since 2018 NAIP imagery was not available for Minnesota to ocularly sample open water.

Post Processing & Validation

We applied the top species model across imagery from each year using GEE to generate a statewide map of the five EAV classes for 2017 and 2018 at a 10 m resolution (Figure 1.4). We applied the water model trained on 2017 samples to imagery from 2017 and 2018 to create an open water mask for each year. The hydrologic feature boundary dataset provided by the MN DNR, was used to mask the terrestrial matrix for each of the four output maps. The binary water/vegetation maps from each year were then applied as a mask to each of the species maps for each year. We implemented a two-step validation using an approach similar to Dennis & LaRoe et al. (*in review*), including both point and polygon overlap metrics (Figure 1.4). First, the top model was applied across validation points that were withheld (Appendix 1B). Secondly, the tabulate area function in ArcGIS (Version 10.4.1) was used to quantify the total area of overlap between the rasterized versions of the original survey polygons (Table 1.1) and model predictions. We also calculated map accuracy confidence intervals based on weighted proportions of area mapped for each class (Olofsson et al., 2014) for the final 2017 map (using 2017 validation data).

Results

Model Results: 2017 and 2018

The accuracy of the species and water models when applied to their respective sets of withheld validation points are shown in Table 1.4. Prior to combining the predictions from the species and water models, their overall accuracies were 89.5 and 95.8 percent correctly classified, respectively. Wild rice had the highest producer's and user's accuracy in the species model (94.9% and 93.3%), while water lilies had the lowest producer's and user's accuracy (81.5% and 84.6%). Water lilies were most commonly confused with cattails, and all other EAV classes were most commonly confused with water lilies. We found trade-offs

between producer's and user's accuracy of the water and vegetation classes in the water model. It was essential to note each model's performance before combining the species and water model predictions because their individual accuracies were factors that influenced the final maps.

Table 1.4. Confusion matrices for the species model (left) and water model (right) including producer's and user's accuracies for each class. The species model was applied to the 8,781 samples (from 2017 and 2018) withheld for validation and had 89.5% overall accuracy. The water model was applied to 5,203 samples withheld (from 2017 only) and had 95.8% overall accuracy. Yellow represents points that were accurately classified, dark green represents user's and producer's accuracies > 90%, and light green represents user's and producer's accuracies > 80%.

Species Model

PREDICTED \ OBSERVED	Wild rice	Water lilies	Rushes	Cattail	Other EAV	Total	User's
Wild rice	1,664	65	25	16	13	1,783	93.3%
Water lilies	29	1,446	56	71	107	1,709	84.6%
Rushes	22	46	1,597	47	43	1,755	91.0%
Cattail	24	140	52	1,606	48	1,870	85.9%
Other EAV	20	58	25	13	1,548	1,644	94.2%
Total	1,759	1,755	1,755	1,753	1,759	8,781	
Producer's	94.9%	81.5%	90.1%	91.1%	88.5%		

Water Model

PREDICTED \ OBSERVED	Water	Vegetation	Total	User's
Water	2,623	90	2,713	96.7%
Vegetation	127	2,363	2,490	94.9%
Total	2,750	2,453	5,203	
Producer's	95.3%	96.3%		

After combining the species and water models' predictions from each year, we found the area of overlap between model predictions and field surveys to be a powerful validation tool for capturing the full scope of model performance, instead solely relying on a small sample of points (Appendix 1B; Table 1.5, Dennis & LaRoe et al., *in review*). Overall, the final maps had decent performance, with some variations in performance between classes and years (Table 1.5). Both maps had similar overlap with cattails, but differed dramatically in their agreement with other EAV. The polygon area utilized for validation excluded any areas that fell beyond the boundaries of hydrologic features, and these areas were often delineated as cattail dominant stands in the field surveys. Prior to masking these areas, the model predictions had high user's and producer's accuracy with the cattails class (Table 1.4). In general, the 2017 final map had greater rates of overlap, and the highest rate of overlap with other EAV (90.1%), rushes (84.1%), and wild rice (90.4%). The 2018 final map had greater overlap with cattails (76.4%) and water lilies (83.9%).

Table 1.5. Percentage of overlap between 2017 and 2018 final maps and rasterized field survey polygons from more homogeneous EAV stands (covering ≥ 60% of a 10 m Sentinel pixel).

EAV Class	2017		2018	
	Polygon Area (m²)	% Overlap	Polygon Area (m²)	% Overlap
Cattails	720,400	72.8	591,300	76.4
Other EAV	861,400	90.1	637,500	54.5
Rushes	1,173,800	84.1	4,991,200	75.5
Water lilies	2,197,500	75.7	3,059,200	83.9
Wild rice	295,900	90.4	659,100	81.7

Map Validation 2017

The 2017 final map (species and water model outputs combined) was validated using weighted proportional area of each class and 2017 validation points from both models (Table 1.6), and this map most accurately captured the wild rice, other EAV, and water classes (Table 1.6). Rushes, other EAV, and wild rice were most commonly misclassified as open water, and open water was most frequently misclassified as rushes (Table 1.6). Water lilies were most frequently misclassified as cattails or wild rice, and cattails were still most frequently misclassified as water lilies and rushes.

Table 1.6. Confusion matrix for final 2017 map (combining species and water models) including mapped area and weights utilizing the framework provided by Olofsson et al. (2014).

		Reference Class						Total Mapped Area (km²)	Weights
		Cattail	Other EAV	Rushes	Water lilies	Wild rice	Water		
Map Class	Cattail	166	12	16	32	8	2	2978.1825	0.211888
	Other EAV	4	764	4	18	1	21	716.807	0.050999
	Rushes	18	2	688	23	7	40	434.2448	0.030895
	Waterlilies	20	11	18	624	16	17	899.7138	0.064012
	Wild Rice	7	8	12	45	714	1	393.4372	0.027992
	Water	4	32	24	29	23	1633	8633.0636	0.614215
	Total	219	829	762	771	769	1714	14055.4489	

We applied proportional, area-based weights to derive confidence intervals that more finely delineated class accuracies which are represented in Table 1.7 (Olofsson et al., 2014). The overall accuracy of the final map was 89.9% ± 0.7% and producer's accuracies for all classes ranged between 75.8% and 95.3%, indicating that the model predictions effectively detected EAV classes and open water once they were

combined. Open water was predicted to be the dominant cover class across lakes and had the highest producer's and user's accuracy. Although the application of the water mask to the species maps increased some of the errors within EAV classes, we observed improvements in visual map assessments compared to those prior to the water masking step. While the producer's accuracy of the rushes class was high (90.3% ± 2%), the user's accuracy was less than half this rate (0.4504 ± 0.0354), which indicates that the results from this class are likely the least reliable due to its frequent misclassification with open water. The metrics for all other classes fell between 98.3% - 45% and demonstrated trade-offs between user's and producer's accuracy. Producer's accuracy for each EAV class remained relatively high, similar to their respective overlap with the original survey polygons (Table 1.4; Table 1.7).

Table 1.7. Confusion matrix for the final 2017 map (species and water models combined) where the accuracy is weighted by proportion of mapped area within each class (Table 1.6). Confidence intervals (95%) are also provided for the overall accuracy, as well as the user's and producer's accuracy for each class based on the framework provided by Olofsson et al., (2014).

	Reference Class										
Map Class		Cattail	Other EAV	Rushes	Water lilies	Wild rice	Water	Total	User's Accuracy	Producer's Accuracy	Overall Accuracy
	Cattail	0.1606	0.0007	0.000649	0.0027	0.0003	0.0007	0.1657	0.9695±0.0228	0.758±0.0187	0.8985±0.0068
	Other EAV	0.0039	0.047	0.000162	0.0015	0.00004	0.0075	0.0601	0.7822±0.0281	0.9216±0.0218	
	Rushes	0.0174	0.0001	0.027895	0.0019	0.0003	0.0143	0.0619	0.4504±0.0354	0.9029±0.0196	
	Water lilies	0.0194	0.0007	0.000730	0.0518	0.0006	0.0061	0.0792	0.6538±0.0336	0.8093±0.0248	
	Wild Rice	0.0068	0.0005	0.000487	0.0037	0.026	0.0004	0.0378	0.6869±0.0328	0.9285±0.0169	
	Water	0.0039	0.002	0.000973	0.0024	0.0008	0.5852	0.5952	0.9831±0.0061	0.9527±0.0217	

The 2018 final map was not validated using these methods since NAIP imagery was not available for Minnesota in 2018, and therefore we did not collect water points in 2018. Additionally, previous literature has highlighted the annual spatial variability of wild rice stands (Rickman et al., 2017) as well as other annual stands of EAV. Therefore, incorporating the water validation samples from the previous year is not appropriate and falls beyond the scope of the 2018 final map. Further validation will be necessary in order to precisely determine the accuracy of open water predictions within the 2018 final map and explore the scope of model applicability through time.

Visual quality assessments provided supplemental details about model performance in specific regions that were difficult to glean from the validation metrics alone (Figure 1.5; Figure 1.6). In general, models were most accurate across more homogenous stands similar to the training data utilized which is well

represented by Big Birch and Little Birch Lakes in 2017 (Figure 1.5) as well as Itasca Lake in 2018 (Figure 1.6). Model misclassification can be visualized within some of the highly mixed EAV stands (particularly the "water lilies and others" class), where the dominant cover type surveyed may be represented in the model predictions, but not predicted as the dominant cover type. These misclassifications are shown across Buffalo Lake in 2017 (Figure 1.5) and Eighth Crow Wing Lake in 2018 (Figure 1.6).

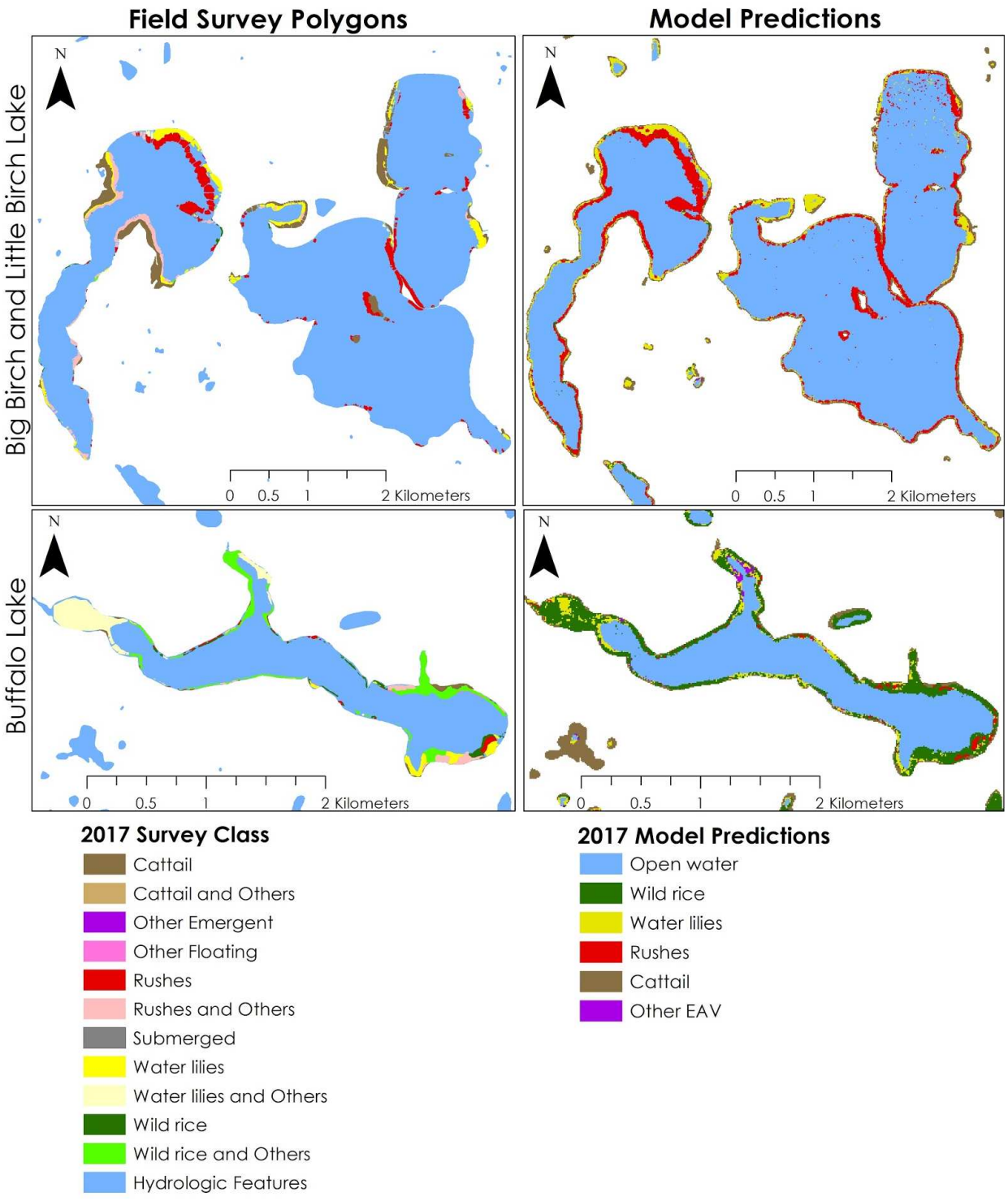


Figure 1.5. Comparison between 2017 survey polygons and 2017 model predictions across Big Birch, Little Birch, and Buffalo Lakes in Minnesota. Note that all classes containing "and Others" contain more heterogeneous EAV stands with greater variation in taxa present; reference Table 1.1 for survey polygon definitions. Additionally, model predictions were constrained to lake boundary polygons; reference Field Survey Data for post processing details and constraints.

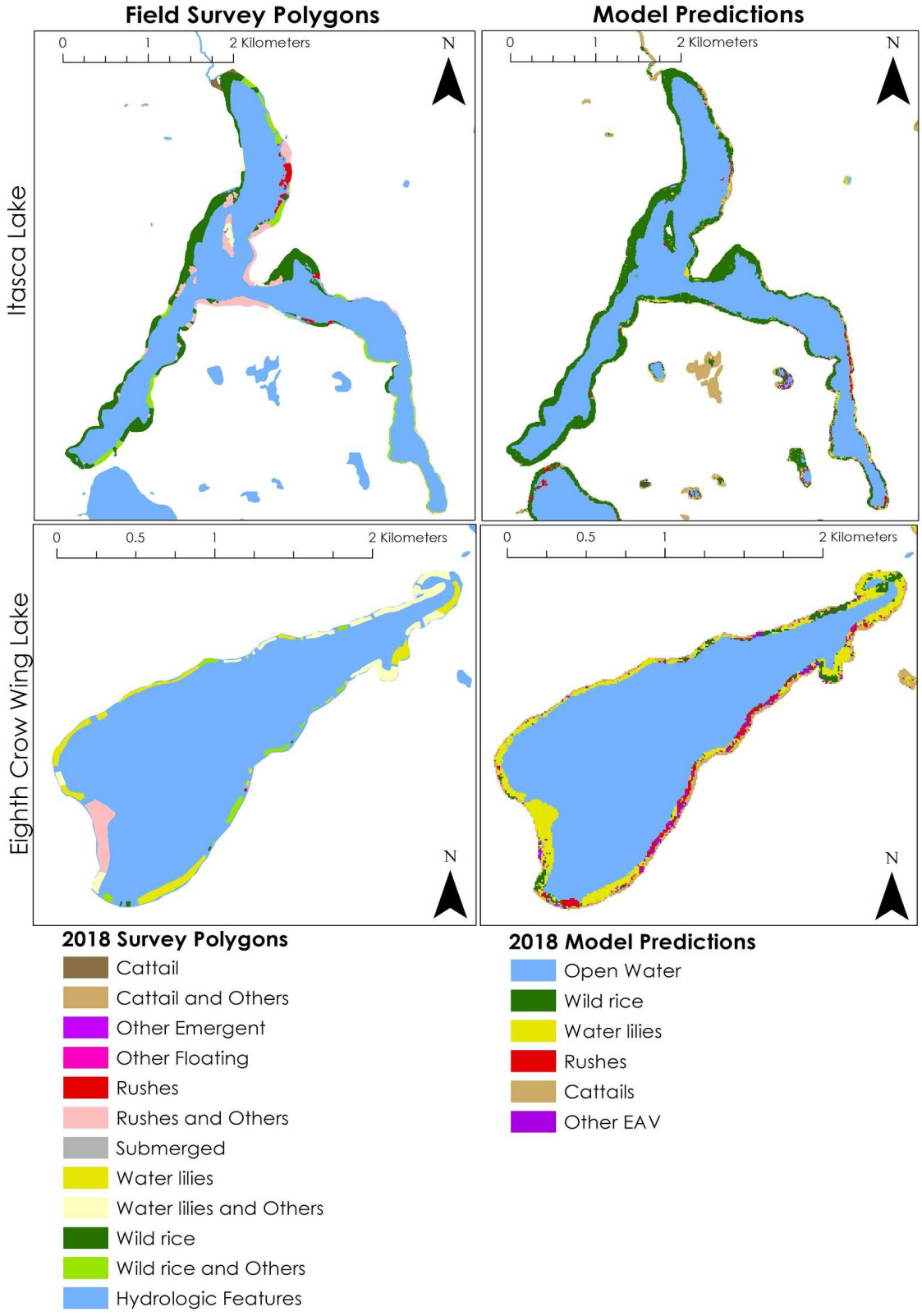


Figure 1.6. Comparison between 2018 survey polygons and 2018 model predictions across Itasca and Eighth Crow Wing Lakes in Minnesota.

Predictors

The random forest species model showed that the normalized difference vegetation index (NDVI) from T2 (June – late July) had the greatest explanatory power for classifying EAV species groups out of 22 variables included in the model, corresponding to seasonal near peak vegetation growth (Table 1.8). The coastal band at various time periods ranked as the next three most important variables. SAR predictors from Sentinel-1 represented 12 out of 22 variables included in the model, and the other 10 variables were derived from Sentinel-2 MSI. Additional predictors utilized in the species model include the normalized difference water index (NDWI), red edge #4 NDVI, red edge #3, and the water vapor band. In total, 10 of these predictors utilized information from T1 (March – late May) and T2, nine predictors utilized information from T3 (August – late September), and only five predictors utilized information from T4 (late September – mid November). The majority of explanatory power was captured from March through late September.

Table 1.8. The final predictors included in the species model with their associated time period and importance ranking in the random forest.

Importance Rank	Mean Decrease Accuracy	Predictor	T1	T2	T3	T4
1	138.168	NDVI		X		
2	131.925	B1 Coastal		X		
3	124.072	B1 Coastal	X			X
4	123.982	B1 Coastal	X			
5	118.163	NDWI	X			
6	111.03	B1 Coastal			X	
7	106.594	B9 Water Vapor	X		X	
8	103.527	B8A Red Edge #4 NDVI	X			
9	100.402	B9 Water Vapor		X	X	
10	95.521	VV range	X		X	
11	95.449	VV range		X	X	
12	95.288	VV median				X
13	90.946	B7 Red Edge #3	X			
14	89.463	VV range		X		X
15	87.706	VV median			X	
16	85.104	VV range	X	X		
17	82.877	VV range			X	X
18	78.872	VH range		X		X
19	77.137	VV median		X		
20	72.627	VH range	X	X		
21	69.059	VH range		X	X	
22	66.898	VH range	X		X	

The water model was created utilizing predictors only from T3 to match the timing of the NAIP imagery used to identify reference points. The five predictors selected in order of rank were VV median, NDVI, the blue band, NDVI red edge #4, and the water vapor band (Appendix 1D). VV median provides a relatively constant backscatter value for open water, and was useful for identifying EAV canopies in the species model. Similarly, NDVI was the top predictor in the species model and is useful for identifying vegetation. VV median may have helped with separating EAV and water specifically, and reduce potential sources of error; the use of NDVI alone would likely cause submersed vegetation or algal blooms to be identified as EAV. Figure 1.7 shows the change in VV range values for each class of EAV between sequential time periods throughout the growing season. Cattails tend to retain a relatively constant median value in each instance based on their perennial structures, and their signature was most distinct utilizing VV range T1-T2 or other T1 variables (Appendix 1C). Rushes were most distinguishable from all classes of EAV based on their distribution of values at NDVI T2, but they tended to have more similar values with at least one other EAV class across other predictors (Appendix 1C). Water lilies and wild rice were a common source of misclassification (Table 1.5; Table 1.6). In Figure 1.7, water lilies and wild rice were only more distinguishable utilizing VV range T2-T3, but they tend to have a relatively similar distribution of spectral and SAR values across many of the predictors considered (Appendix 1C).

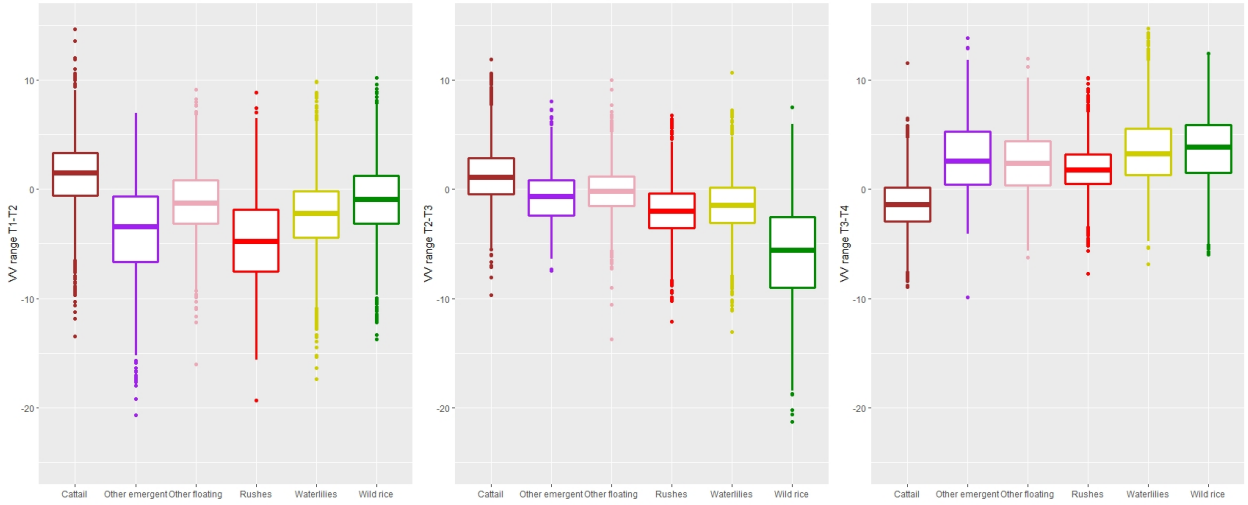


Figure 1.7. Distribution of VV range variable values by EAV class (Brown = cattails, purple = other emergent, pink = other floating, red = rushes, yellow = water lilies, green = wild rice across the growing season (Left: T1 – T2, Middle: T2 – T3, Right: T3 – T4).

Discussion

The methods and promising results outlined in this study indicate a strong potential for creating annual EAV maps across Minnesota to support conservation and proactive management. By employing robust field survey data in conjunction with remote sensing techniques that were specifically refined to capture species phenology, we achieved high rates of overlap between model predictions and the field survey polygons. The results demonstrate that it is feasible to accurately separate classes of EAV at a spatial resolution relevant for land management at broad spatial extents. Our classification accuracies were comparable to previous studies despite having a greater number of distinctive EAV classes (Sawaya et al., 2003; Price, 2012; Dennis & LaRoe et al.., *in review*). The cumulative importance of all VV range variables included in the species model were not directly compared with the results of Dennis & LaRoe et al. (*in review*) due to the differing target classes of EAV. However, our model predictions yielded higher overlap with wild rice survey polygons, which may suggest that the additional phenological time periods utilized and/or methods for selecting training samples from more homogeneous EAV stands improved wild rice detection (Table 1.4; Figure 1.6).

Previous literature has demonstrated the capacity for NDVI to improve vegetation classification accuracies (Sawaya et al., 2003; Peñuelas et al., 1993) and that the Coastal band from Sentinel-2 is useful for mapping marine habitats and bathymetry (Immordino et al., 2019; Poursanidis et al., 2019). NDVI may have captured the bulk of individual vegetation class characteristics, and the repeated significance of the Coastal band at various time periods (Table 1.8) may suggest that water level variability as well as water quality are two environmental factors contributing to the variance explained in the model. There was a nearly even split between predictors utilized from both Sentinel-1 SAR and Sentinel-2 MSI, which may indicate that EAV classification improvements stemmed from the combination of spectral and SAR variables. The final predictors in the species model utilized information that spanned almost evenly across the first three time periods, and of these 17 predictors, nine of them utilized the range of values between two different time periods (Table 1.8). Only five out of 22 predictors utilized information from the fourth time period (T4, late

September to mid-November), and four of these predictors were based on the range of values between another time period and T4 (Table 1.8). Our results suggest that the novel use of four phenologically distinct time periods during seasonal EAV growth captured a finer level of detail within each EAV class through time which may have improved spectral separability between the classes (Appendix 1C).

There were several challenges we faced while generating statewide maps of EAV based on the computational and data processing requirements. Due to limitations within GEE, it was not possible to incorporate more than approximately 23,000 training points (total) within the random forest model. Creating two separate models, one to better distinguish between species and one to separate water, maximized the number of training samples utilized within each class to provide more robust predictions covering the full extent of Minnesota. While producer's and user's accuracy of each EAV class was lower after applying the water mask to the species model (Table 1.4; Table 1.7), this approach still provided higher accuracy rates for individual classes compared to utilizing a single predictive model.

While our models showed promising results for most EAV classes, water lilies were most commonly confused with other species in the model It is possible that the SAR variables struggled to capture backscatter signals from water lilies that were distinct from open water. In mixed heterogeneous EAV stands, water lilies were often confused with wild rice (Figure 1.6) and there are potential explanations for this observation aside from the frequent co-occurrence of the two. Between May and June (end of T1 and start of T2), wild rice enters the floating leaf stage where it tends to have more similar spectral and backscatter responses to those of water lilies floating on the water's surface (Appendix 1C). During emergence in mid June through July, wild rice has tall vertical structures that would provide a canopy overhead of nearby water lilies. At peak growth, wild rice may obstruct or alter the reflectance and backscatter values of water lilies.

Additionally, the confusion between open water and rushes in both years may have resulted from the considerably low NDVI values of rushes. At NDVI T2, the mean value of the rushes class is approximately zero, even though the distribution of these values is large (Appendix 1C). The species of EAV, total area, and location of the field surveys for the other EAV class (originally surveyed as other floating and other emergent) differed dramatically between 2017 and 2018 (Table 1.4) which may account for the large

differences in their overlap with EAV survey polygons (90% and 54%). Additionally, a greater number of training samples were included from 2017 based on a small sample size of other EAV in 2018. There were few remaining validation points for the cattails class within the hydrologic boundaries (Table 1.4) considering that they comprised 21% of the total predicted area (Table 1.6). Further independent validation would help clarify the discrepancies between their user's and producer's accuracy for the final 2017 map. In future efforts to map cattails utilizing remotely sensed data, it may be necessary to incorporate additional samples (field surveys or ocular samples) representative of forests or other wetland vegetation in order to distinguish between these species.

While the water model performed well when applied to 2018, it was only trained using samples from 2017 which may have resulted in the lower overlap agreements in 2018 (Table 1.4). Ocular samples or field surveys containing open water locations each year would likely provide the greatest improvement. These results may suggest that a model based on a single year of data cannot retain the same degree of accuracy when applied through time without adding training samples from the year of interest. Moreover, this also indicates that it is critical for detailed annual field survey data to be continuously collected and utilized for finely tuning annual EAV distribution models, although mapping resources may aid in strategic field sampling efforts and a targeted stratified sampling design based off of the previous year predictions. Alternatively, there may be a threshold at which a single model may be applicable once several years of field surveys are incorporated; however, the results of this study do not provide any indication of this threshold or the number of survey data years required to create a single model with high temporal accuracy due to our limited 2-year study time period.

In light of these results, the presented statewide EAV maps have strong implications for local and landscape management and conservation efforts due to their fine spatial resolution and continuous coverage across Minnesota water bodies. The detailed methods outlined and predictor variables that we utilized can be used to generate additional maps by incorporating Sentinel-1 and Sentinel-2 imagery with EAV species locations and open water samples from the year of interest. The broad spatial extent of these maps can provide supplemental years of data for lakes since it is not feasible to survey every lake each year. Since it is

not possible to sample exceptionally dense wild rice stands that cover some of the major harvesting lakes in Minnesota, annual EAV maps have the potential to estimate cover in these instances. They may also provide new information about lakes and rivers that have not been previously surveyed based on their locations or challenging terrain that may not be safe to navigate. Furthermore, these maps may have utility for identifying additional lakes with unique EAV community compositions that are of high conservation interest, in addition to previously established sentinel lakes and other prioritized areas. They may also reveal the tradeoffs between different conservation management and restoration techniques and help with prioritizing efforts. The spatial predictions of EAV may also be important for assessing landscape drivers of their communities from a holistic point of view.

If statewide EAV maps are created for additional years, they could potentially be used to evaluate temporal dynamics in EAV communities and quantify changes in percent cover. Maps could be useful for identifying lakes that are changing more rapidly than others, or highlight areas of potential concern where it may be important to conduct field surveys the following year. Furthermore, this information would significantly benefit invasive species management efforts, as maps spanning a broader period of time could provide opportunities to recognize the onset of cattail encroachment and target management efforts within a shorter time frame. They may also reveal locations vulnerable to invasion where the loss of wild rice or other native species may be rapidly shifting community composition, supporting more efficient and proactive management measures to address changes.

Conclusion

The combination of detailed field surveys paired with Sentinel-1 SAR and Sentinel-2 MSI at four phenologically significant time periods provided EAV class predictions across Minnesota with reasonable accuracy in 2017 and 2018. All maps were generated at a spatial resolution relevant to land management, and this study has potential to bring forth landscape level trends not previously recognized. These methods set the stage for subsequent analyses at broader spatial scales to quantify temporal shifts or trends in EAV

communities, as well as evaluate potential landscape and environmental drivers of species distributions. The combination of these diverse and detailed datasets provides promising methods for generating statewide, annual maps of EAV distribution and ultimately provide the first step toward approaching landscape-scale conservation of EAV communities in Minnesota.

REFERENCES

Alahuhta, J. (2014). Geographic patterns of lake macrophyte communities and species richness at regional scale. Journal of Vegetation Science, 26(3), 564–575. Doi: 10.1111/jvs.12261

Apfelbaum, S. I. (1985). Cattail (Typha spp.) Management. Natural Areas Journal, 5(3), 9–17. Retrieved from https://www.jstor.org/stable/43910841

Biesboer, D. D. (2019). The Ecology and Conservation of Wild Rice, *Zizania palustris L.*, in North America. *Acta Limnologica Brasiliensia*, *31*(102). Doi: 10.1590/s2179-975x2319

Blann, Kristen L., Anderson, James L., Sands, Gary R., & Vondracek, Bruce (2009). Effects of Agricultural Drainage on Aquatic Ecosystems: A Review, Critical Reviews in Environmental Science and Technology, 39:11, 909-1001, DOI: 10.1080/10643380801977966

Bourgeau-Chavez, L., Riordan, K., Powell, R. B., Miller, N., & Nowels, M. (2009). Improving Wetland Characterization with Multi-Sensor, Multi-Temporal SAR and Optical/Infrared Data Fusion. Advances in Geoscience and Remote Sensing. Doi: 10.5772/8327

Breiman, L. (2001). Random forests. Machine Learning, 45, 5–32.

Brisco, B., Li, K., Tedford, B., Charbonneau, F., Yun, S., & Murnaghan, K. (2013) Compact polarimetry assessment for rice and wetland mapping, International Journal of Remote Sensing, 34:6, 1949-1964, DOI: 10.1080/01431161.2012.730156

Dabboor, M., & Brisco, B. (2019). Wetland Monitoring and Mapping Using Synthetic Aperture Radar. Wetlands Management – Assessing Risk and Sustainable Solutions. Doi: 10.5772/intechopen.80224

Dennis, K. *, LaRoe, J. *, Vorster, A., Young, N. E., Evangelista, P. E., Mayer, T., Carver, D., Simonson, E., Martín, V., Kern, A., Radomski, P., Knopik, J., & Khoury, C. K. (2020). Improved Remote Sensing Methods to Detect Northern Wild Rice (*Zizania palustris L.*). Remote Sensing. Manuscript submitted for publication.

Dubbe, D., Garver, E., & Pratt, D. (1988). Production of cattail (Typha spp.) biomass in Minnesota, USA. Biomass, 17(2), 79–104. Doi: 10.1016/0144-4565(88)90073-x

Evans JS, Murphy MA, Holden ZA, & Cushman SA (2011). "Modeling species distribution and change using Random Forests." In Drew CA, Wiersma YF, Huettmann F (eds.), _Predictive species and habitat modeling in landscape ecology: concepts and applications_, chapter 8, 139-159. Springer, New York. ISBN 978-1-4419-7390-0.

Forman, R. T. T., & Gutzwiller, K. J. (2002). Applying landscape ecology in biological conservation. New York: Springer.

Gallant, A.L., Kara, S.G., White, L., Brisco, B., Roth, M.F., Sadinski, W., & Rover, J. (2014). Detecting Emergence, Growth, and Senescence of Wetland Vegetation with Polarimetric Synthetic Aperture Radar (SAR) Data. Water, 6(3), 694-722.

Gonoski, J., Burk, T. E., Bolstad, P. V., & Balogh, M. (2005). Rice Lake National Wildlife Refuge Historic Wild Rice Mapping (1983 – 2004). College of Natural Resources and Minnesota Agricultural Experiment Station University of Minnesota.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment, 202, 18-27.

Hansel-Welch, N., Butler, M. G., Carlson, T. J., & Hanson, M. A. (2003). Changes in macrophyte community structure in Lake Christina (Minnesota), a large shallow lake, following biomanipulation. Aquatic Botany, 75(4), 323–337. Doi: 10.1016/s0304-3770(03)00002-0

Immordino, F., Barsanti, M., Candigliota, E., Cocito, S., Delbono, I., & Peirano, A. (2019). Application of Sentinel-2 Multispectral Data for Habitat Mapping of Pacific Islands: Palau Republic (Micronesia, Pacific Ocean). Journal of Marine Science and Engineering, 7(9), 316. Doi: 10.3390/jmse7090316

Joniak, T., Kuczyńska-Kippen, N., & Nagengast, B. (2007). The role of aquatic macrophytes in microhabitatual transformation of physical-chemical features of small water bodies. Hydrobiologia, 584(1), 101–109. Doi: 10.1007/s10750-007-0595-8

Kissoon, L. T. T., Jacob, D. L., Hanson, M. A., Herwig, B. R., Bowe, S. E., & Otte, M. L. (2013). Macrophytes in shallow lakes: Relationships with water, sediment and watershed characteristics. Aquatic Botany, 109, 39–48. Doi: 10.1016/j.aquabot.2013.04.001

Larkin, D. J., Freyman, M. J., Lishawa, S. C., Geddes, P., & Tuchman, N. C. (2011). Mechanisms of dominance by the invasive hybrid cattail Typha × glauca. Biological Invasions, 14(1), 65–77. Doi: 10.1007/s10530-011-0059-y

Mansaray, L. R., Huang, W., Zhang, D., Huang, J., & Li J. (2017). Mapping rice fields in urban Shanghai, southeast China, using Sentinel-1A and Landsat-8 datasets. Remote Sensing, 9 (3), p. 257.

MNDR.(n.d) Minnesota Climate Trends 1895-2018. Retrieved from https://arcgis.dnr.state.mn.us/ewr/climatetrends/

Moyle, J. (1956). Relationships between the Chemistry of Minnesota Surface Waters and Wildlife Management. The Journal of Wildlife Management, 20(3), 303-320. Doi:10.2307/3796967

Muthukrishnan, R., & Larkin, D. J. (2020). Invasive species and biotic homogenization in temperate aquatic plant communities. Global Ecology and Biogeography, 29(4), 656–667. Doi: 10.1111/geb.13053

Myrbo, A., Swain, E. B., Engstrom, D. R., Wasik, J. C., Brenner, J., Shore, M. D., Peters, E. P., Blaha, G. (2017) (a). Sulfide Generated by Sulfate Reduction is a Primary Controller of the Occurrence of Wild Rice (Zizania palustris) in Shallow Aquatic Ecosystems. Journal of Geophysical Research: Biogeosciences, 122(11), 2736-2753. Doi:10.1002/2017jg003787

Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., & Wulder, M. A. (2014). Good practices for estimating area and assessing accuracy of land change. Remote Sensing of Environment, 148, 42–57. Doi: 10.1016/j.rse.2014.02.015

Peñuelas, J., Gamon, J. A., Griffin, K. L., & Field, C. B. (1993). Assessing community type, plant biomass, pigment composition, and photosynthetic efficiency of aquatic vegetation from spectral reflectance. Remote Sensing of Environment, 46(2), 110–118. Doi: 10.1016/0034-4257(93)90088-f

Pillsbury, R. W., & Mcguire, M. A. (2009). Factors affecting the distribution of wild rice (*Zizania palustris*) and the associated macrophyte community. Wetlands, 29(2), 724–734. Doi: 10.1672/08-41.1

Poursanidis, D., Traganos, D., Reinartz, P., & Chrysoulakis, N. (2019). On the use of Sentinel-2 for coastal habitat mapping and satellite-derived bathymetry estimation using downscaled coastal aerosol band. International Journal of Applied Earth Observation and Geoinformation, 80, 58–70. Doi: 10.1016/j.jag.2019.03.012

Price, M. W. (2012). Spectral identification of wild rice (*Zizania palustris* L.) using indigenous knowledge and Landsat multispectral data (Master's Thesis). Retrieved from the ScholarWorks at University of Montana Graduate Student Theses, Dissertations, & Professional Papers database. https://scholarworks.umt.edu/etd/910

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.URL http://www.R-project.org/.

Radomski, P., & Goeman, T. J. (2001). Consequences of Human Lakeshore Development on Emergent and Floating-Leaf Vegetation Abundance. North American Journal of Fisheries Management, 21(1), 46–61. Doi: 10.1577/1548-8675(2001)021<0046:cohldo>2.0.co;2

Radomski, P. (2006). Historical Changes in Abundance of Floating-Leaf and Emergent Vegetation in Minnesota Lakes. North American Journal of Fisheries Management, 26(4), 932–940. Doi: 10.1577/m05-085.1

Radomski, Paul & Woizeschke, Kevin & Carlson, Kristin & Perleberg, Donna. (2011). Reproducibility of Emergent Plant Mapping on Lakes. North American Journal of Fisheries Management. 31. 144-150. 10.1080/15222055.2011.562744.

Rahel, F. J., & Olden, J. D. (2008). Assessing the Effects of Climate Change on Aquatic Invasive Species. Conservation Biology, 22(3), 521–533. Doi: 10.1111/j.1523-1739.2008.00950.x

Reinke, K., & Jones, S. (2006). Integrating vegetation field surveys with remotely sensed data. Ecological Management and Restoration, 7(s1). Doi: 10.1111/j.1442-8903.2006.00287.x

Rickman, D. L., Greensky, W. A., Al-Hamdan, M. Z., Estes, M. G., Crosson, W. L., & Estes, S. M. (2017). Spatial and Temporal Analyses of Environmental Effects on *Zizania palustris* and Its Natural Cycles. NASA Technical Reports Server. Retrieved from https://ntrs.nasa.gov/search.jsp?R=20170010653

Rundquist, D. C., Narumalani, S., & Narayanan, R. M. (2001). A review of wetlands remote sensing and defining new considerations. Remote Sensing Reviews, 20(3), 207–226. Doi: 10.1080/02757250109532435

Sawaya, K. E., Olmanson, L. G., Heinert, N. J., Brezonik, P. L., & Bauer, M. E. (2003). Extending satellite remote sensing to local scales: land and water resource monitoring using high-resolution imagery. Remote Sensing of Environment, 88(1-2), 144–156. Doi: 10.1016/j.rse.2003.04.006

Silva, T. S. F., Costa, M. P. F., Melack, J. M., & Novo, E. M. L. M. (2007). Remote sensing of aquatic vegetation: theory and applications. Environmental Monitoring and Assessment, 140(1-3), 131–145. Doi: 10.1007/s10661-007-9855-3

Villa, P., Pinardi, M., Bolpagni, R., Gillier, J.-M., Zinke, P., Nedelcuţ, F., & Bresciani, M. (2018). Assessing macrophyte seasonal dynamics using dense time series of medium resolution satellite data. Remote Sensing of Environment, 216, 230–244. Doi: 10.1016/j.rse.2018.06.048

Wilcox, D. A., & Meeker, J. E. (1992). Implications for faunal habitat related to altered macrophyte structure in regulated lakes in northern Minnesota. *Wetlands*, 12(3), 192–203. Doi: 10.1007/bf03160609

Wilcox, D. A., Kowalski, K. P., Hoare, H. L., Carlson, M. L., & Morgan, H. N. (2008). Cattail Invasion of Sedge/Grass Meadows in Lake Ontario: Photointerpretation Analysis of Sixteen Wetlands over Five Decades. Journal of Great Lakes Research, 34(2), 301–323. Doi: 10.3394/0380-1330(2008)34[301:ciogmi]2.0.co;2

Young NE, Jarnevich CS, Sofaer HR, Pearse I, Sullivan J, Engelstad P, et al. (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

CHAPTER 2: RELATIVE INFLUENCE OF SYSTEMIC DRIVERS AND THEIR RELATIONSHIPS WITH NORTHERN WILD RICE

Introduction

Significance of Wild Rice

Northern wild rice (*Zizania palustris L.*), or wild rice, is an indicator species found growing annually within shallow water in Minnesota, and its populations have started to disappear throughout the last century (Pillsbury & McGuire, 2009). It is an essential, multifaceted resource serving many ecological, cultural, and economic roles, making it a high conservation priority (Chapter 1, section 1.2; Biesober, 2019). Wild rice provides food and habitat to many species of fish and waterfowl (McAtee, 1917), many of which are prioritized for conservation (DNR, 2008). Traditionally known as *manoomin,* wild rice is an important cultural resource to indigenous peoples of Minnesota who have harvested this grain for over two millennia (Biesober, 2019). Cultivated stands of wild rice also serve as an important agricultural and economic resource in Minnesota (Kennard et al., 1999), and it is known as a crop wild relative that has the potential to improve crop resistance to pests, drought, and other threats (Khoury et al., 2013).

Previous studies have identified fundamental drivers of wild rice, including hydrologic flow regimes (Blann et al., 2009; DNR, 2008), water biogeochemistry, water transparency (Myrbo et al., 2017(a); Myrbo et al., 2017(b); Pastor et al., 2017; Pollman et al., 2017), temperature, and anthropogenic land-use change or other disturbance (Muthukrishnan & Larkin, 2020; Pillsbury & McGuire, 2009; DNR, 2008). The myriad of influences on wild rice make it challenging to untangle the relationships between systemic drivers and quantify the importance of each component, such as biological processes of dispersal and habitat preferences. Previous studies have not examined systematic drivers alongside continuous wild rice cover across HUC08 watershed or broader regions because the distribution of wild rice on many lakes and rivers is unknown (DNR, 2008). In this study, we utilized recent advances in the mapping of wild rice across Minnesota (Chapter 1) to investigate the directional relationships between wild rice and key environmental influences to

identify significant drivers with potential management and conservation implications for the preservation of native wild rice stands.

Drivers and Scale

Previous studies have recognized that the scales selected for observing and investigating ecological questions affect the observed patterns, and thus, the inferences drawn (Levin, 1992). Observed ecological patterns are often the result of top-down (e.g. climate and surrounding land cover/land use) and bottom-up drivers (e.g. biological controls such as dispersal mechanisms) with potential interactions between these scales. These principles extend to wild rice ecology, where drivers of observed distribution patterns may not be evident without examining relationships across continuous landscapes and by incorporating a variety of spatial scales.

Wild rice is influenced by a variety of top-down drivers and regional patterns that affect its annual distribution. Considering the full geographic extent of suitable habitat, wild rice grows across lakes that serve as patches connected via hydrologic flows which are nestled in the terrestrial matrix. These patches (lakes) containing wild rice are influenced by pollution, sediments, and nutrient inputs from the surrounding terrestrial matrix, and the impacts accumulate and drive observed system dynamics downstream (Blann et al., 2009). Many of the inputs from terrestrial systems into aquatic ecosystems are also influenced by top-down drivers; these include factors related to climate such as precipitation, temperature, and the frequency and duration of extremes (Fischer & Knutti, 2015). Additionally, hydrologic flows are manipulated through dams, diversions, and ditches that alter flow rates, flow volumes, and water transparency which negatively impact wild rice (Blann et al., 2009; DNR, 2008). In the floating leaf stage, wild rice is particularly sensitive to changes in hydrologic flows and dramatic changes in water level (Pillsbury & McGuire, 2009; DNR, 2008). These changes include frequency and intensity of heavy rainfall and flooding events, as well as effects from dams (Pillsbury & McGuire, 2009; DNR, 2008). Dams can cause channelization and increased abundance of sediment accumulation which can bury natural seed banks of wild rice; or in contrast, dams can eliminate variability in flow allowing for the encroachment of shoreline or perennial species (DNR, 2008). However, previous studies have suggested that wild rice thrives with brief, low magnitude, and infrequent natural

disturbances in hydrologic flows that prevent water lilies from outcompeting wild rice, benefiting it long term (Meeker & Tillison, 2018; Myrbo et al., 2017(a); Pillsbury & McGuire, 2009).

In ecosystems with minimal anthropogenic disturbances or nearby agriculture, previous studies identified intraspecific competition and nutrient limitations as factors primarily influencing the abundance of wild rice (Pillsbury & McGuire, 2009; Lee, 2002). In altered ecosystems, surface water sulfate is added primarily by anthropogenic disturbances through mining and agricultural runoff and poses a serious threat to freshwater ecosystems and wetlands (Lamers et al., 2002). Structural equation modeling of wild rice drivers through a random stratified sample of lakes and mesocosm experiments, identified pore water sulfide toxicity and surface water sulfate as top-down drivers of wild rice abundance and persistence (Figure 2.1; Myrbo et al., 2017(a); Pollman et al., 2017). Furthermore, wild rice habitat has been threatened by dams, urbanization or shoreline development (which also degrades water quality), and invasive species (Meeker & Tillison, 2018; Pillsbury & McGuire, 2009; DNR, 2008).

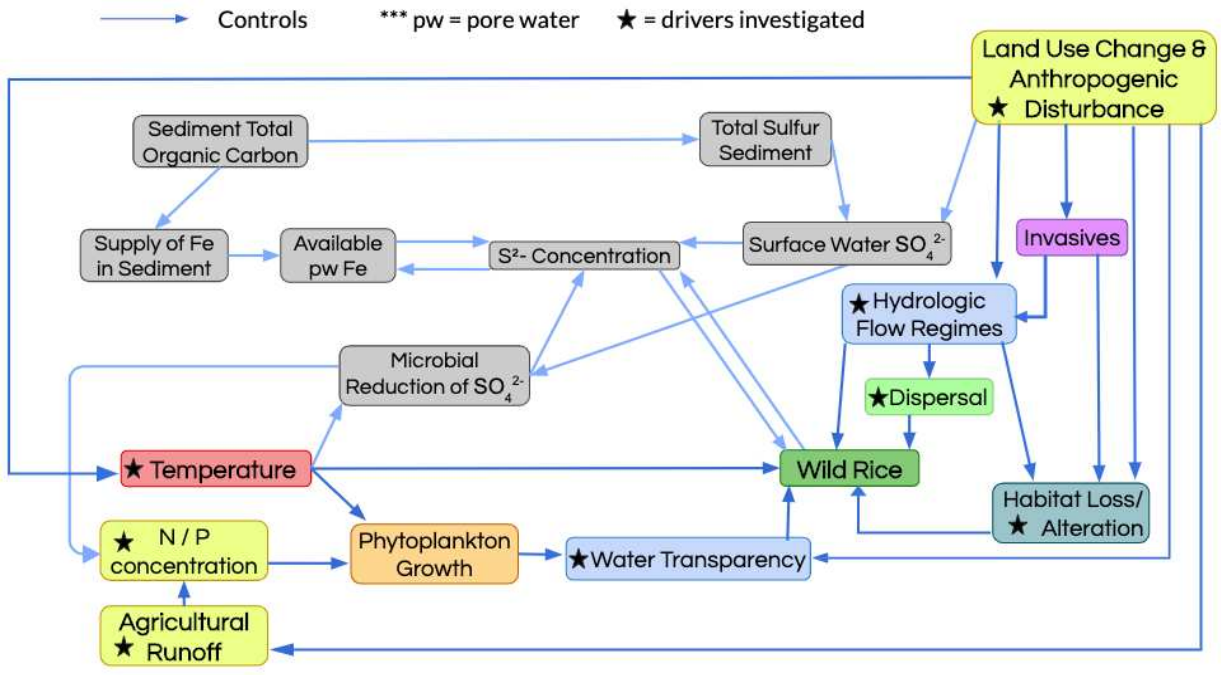


Figure 2.1. Conceptual diagram of drivers and their influences on wild rice. This diagram expands upon structural equation modeling of the primary controls that sulfate/sulfide concentrations have on wild rice (Pollman et al., 2017). Literature synthesis revealed many additional systematic drivers also influencing abundance and distribution of wild rice, and the direction of each influence is represented by a blue arrow. Stars represent drivers considered within this study by either direct or indirect metrics. Note: not all potential arrows are included for the sake of clarity. (N = nitrogen, P = phosphorus, Fe = Iron, S^{2-} =sulfide, SO_4^{2-} = sulfate)

In contrast, the biological traits and mechanisms unique to wild rice may influence its distribution, or act as bottom-up drivers (Gripenberg & Roslin, 2007). Wild rice is found within shallow waters less than 1.5 m in depth, and it primarily relies on wind pollination and sexual reproduction (Myrbo et al., 2017(a); Kahler et al., 2014). Neutral water pH is associated with higher density wild rice stands, and water transparency is crucial for wild rice seed germination which requires sunlight to penetrate the water (Myrbo et al., 2017(a); Pillsbury & McGuire, 2009; Scheffer, 1998). Wild rice is also limited by nutrient availability and winter freezing temperatures, and requires nearly three months of consistent freezing temperatures to break seed dormancy and successfully germinate at a high rate the following year (Myrbo et al., 2017(a); Kahler et al., 2014; Lee, 2002; Kovach & Bradford, 1992).

Although seed dispersal is also thought to be a limiting factor in the distribution or spread of wild rice (DNR, 2008), to our knowledge, no previous studies have explicitly examined dispersal mechanisms or limitations of natural wild rice in the context of its comprehensive distribution. Natural wild rice seeds are dispersed through mechanisms unique to cereal crops known as "shattering" (Kennard, Phillips, & Porter, 2002) where mature seeds will shatter and disperse quickly during thunderstorms or during stronger winds (Kahler et al., 2014). Shattering is a trait that has been selected against when cultivating commercial wild rice to improve crop yields, and non-shattering varieties have been adapted (Kahler et al., 2014). However, even in non-shattering cultivars, a case study found that the introduction of the natural varieties of wild rice (with shattering traits) caused the shattering gene to rapidly proliferate throughout the cultivar over time (Kahler et al., 2014). This demonstrates that shattering is a trait selected for in natural wild rice stands, and leads to greater reproductive success within its natural environment (Vittori et al., 2019). This allows for seeds to be more widely dispersed by waterfowl, wind, and hydrologic flows (Delouche & Bugros, 2007) which suggests that waterfowl, wind, and hydrologic flows are likely key mechanisms driving dispersal. A previous study found that wetland vegetation species dispersing primarily through water were not capable of dispersing as far as species relying on wind (Soomers et al., 2012), and this may imply that wind throughout the short two-week wild rice harvest time period may be more strongly related to dispersal, and ultimately, distributions.

The Framework

To account for the complex suite of potential bottom-up and top-down factors that may influence species distributions, Robledo-Arnuncio et al. (2014) suggests utilizing a mosaic of modeling methods to identify key drivers including: 1) correlational relationship exploration of landscape level drivers and their variance across space, and 2) selecting the key environmental predictors based on the specific habitat niches, morphological traits, and ecology of a species and identifying interactions between these predictors. However, most of these environmental and ecological datasets utilized in this context contain inherent spatial autocorrelation. Methods for dealing with spatial autocorrelation have existed for many years to reduce impacts on parameter coefficients and inferences drawn from models (Miller, Franklin, & Aspinall, 2007; Dormann, 2007; Cressie, 1993). Environmental gradients, competition, and dispersal are primary sources of spatial autocorrelation within species distribution data, with the degree of spatial autocorrelation varying across spatial scales (Dormann, 2007). Spatial autocorrelation is often related to the selected response variable in a model, and it has been found more frequently when summarizing species data across patches on the landscape (Dormann, 2007). Therefore, methods to address spatial autocorrelation are important to consider within a modeling framework to examine drivers of a species' spatial distribution.

Objectives of Research

Within this study, we investigate systematic drivers of wild rice abundance across individual lakes and their relative influences at varying spatial scales, utilizing predicted wild rice maps from 2018 (Chapter 1). Drivers considered in this study (Figure 2.1) are assessed through a variety of spatial, ecological, hydrological, and environmental variables that may directly or indirectly impact wild rice cover, and these variables were summarized at catchment, lake and HUC08 watershed scales. We develop and test a three-stage modeling approach: 1) assess variable correlations with wild rice cover, select variables, and investigate preliminary models using random forest; 2) test for and identify significant interactions between selected variables through multiple linear regressions; and 3) test for spatial autocorrelation and utilize the final selected set of variables and interactions to fit a spatial lag model to account for spatial autocorrelation. The objectives of this study are to test our 3-stage modeling approach for the simultaneous investigation of potential bottom-

up and top-down drivers of wild rice, quantify the directionality of their relationships, and provide initial interpretations of the potential management and conservation implications of our results.

Methods

Study Area

Our study area consisted of nine HUC08 watersheds located in North-central Minnesota (Figure 2.2) within the headwaters of the larger Upper Mississippi River watershed (HUC02). These adjacent watersheds captured wide gradients of elevation, temperature, land cover, and biodiversity. Between 1981 and 2019, the mean annual precipitation across these watersheds was approximately 67.5 cm, while our study year, 2018, had approximately 62 cm of cumulative precipitation, which was 5.5 cm less than the mean annual (MNDR, n.d.). There were 590 lakes within this region that had associated bathymetric data necessary for our analyses; 366 of these lakes contained predicted wild rice presence in 2018 and were selected as the sample lakes for our study. Among our selected study lakes are some that host a variety of long term conservation and management efforts (sentinel lakes).

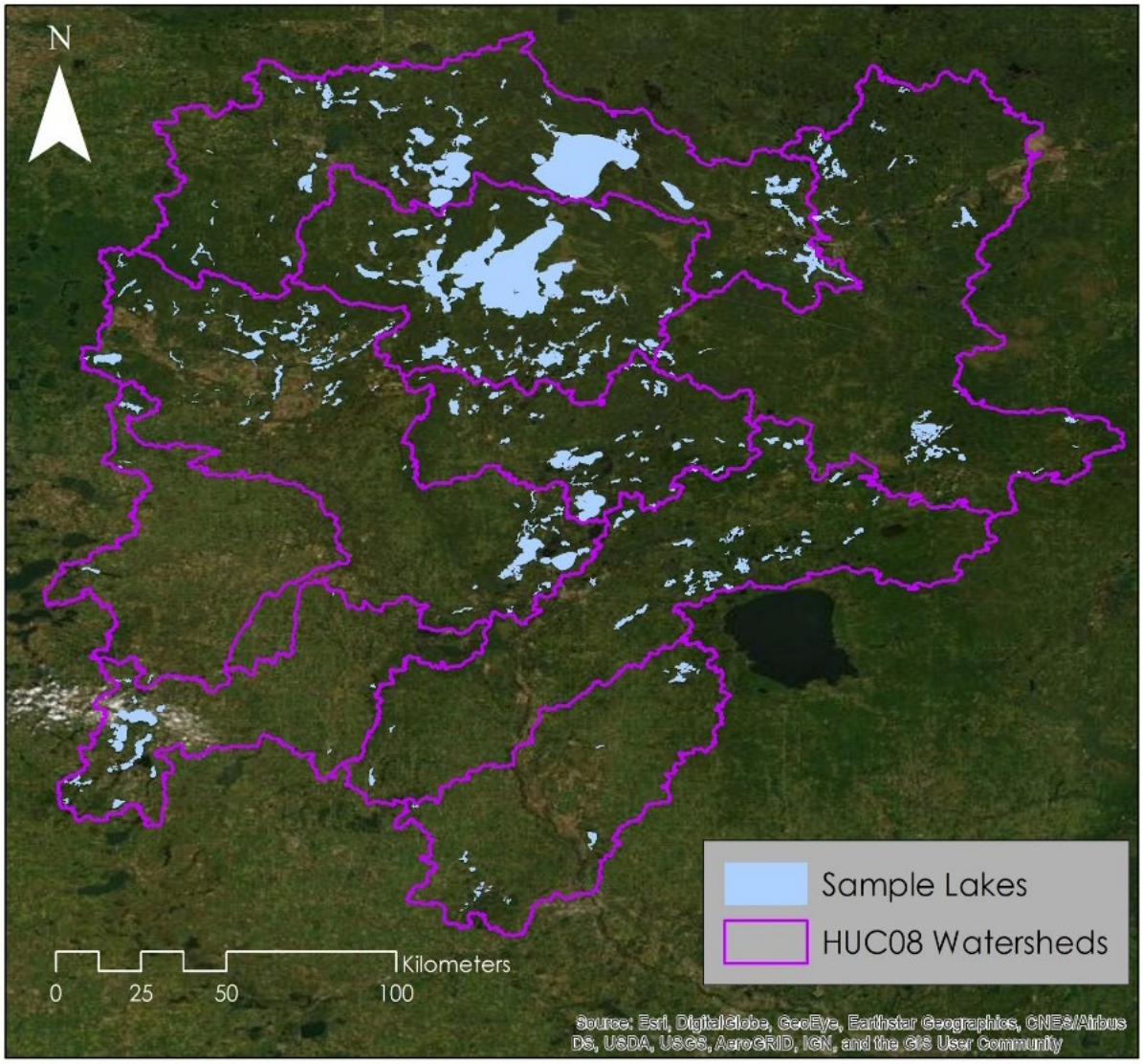


Figure 2.2. The study region is delineated by nine HUC08 watersheds and 366 sample lakes utilized in the analysis. Samples were restricted to lakes that had available bathymetry data as well as predicted wild rice presence in 2018. The watersheds reside within the larger Upper Mississippi River watershed (HUC02) which includes the major headwaters.

Data Preparation

We employed maps of wild rice presence in 2018 at a 10 m spatial resolution (Chapter 1) as the primary data source depicting comprehensive abundance, distribution, and extent across the state. Wild rice presence from this map was isolated to nine HUC08 watersheds. As noted in Chapter 1, the predictive models may have over predicted wild rice in some locations that contained highly heterogeneous stands of emergent aquatic vegetation. To improve the confidence from the model predictions, we utilized a minimum mapping unit of 11 adjacent pixels as a threshold for wild rice presence to reduce potential noise or errors

(Fournier et al., 2007). We standardized wild rice cover on each lake using the proportional values of wild rice cover within areas designated as suitable habitat (water depths less than 1.5 m or 4.9 ft; Price, 2012) using lakes with available bathymetric maps (Minnesota Department of Natural Resources, 2014). It was important to note that water levels and their boundaries may vary across years because the bathymetric contours and lake boundary features were generated from data collected more than a decade prior to the study time period. Considering that cumulative precipitation in 2018 was lower than the mean annual precipitation (1981-2019), and with the restriction of 5 ft intervals for the bathymetric maps, all lake regions with a depth of 3.05 m or less (10 ft) were considered as "potential" suitable habitat (Equation 1).

$$Wild\ Rice\ Cover = \frac{Area\ of\ Wild\ Rice\ m^2}{Lake\ Area\ <3.05m\ depth\ m^2} \qquad \textbf{Equation 1}$$

Initial evaluations of the proportional cover of wild rice revealed a heavily right-skewed distribution, warranting a log transformation to normalize its distribution (Figure 2.3). Log transformed wild rice cover served as the response variable for all subsequent modeling.

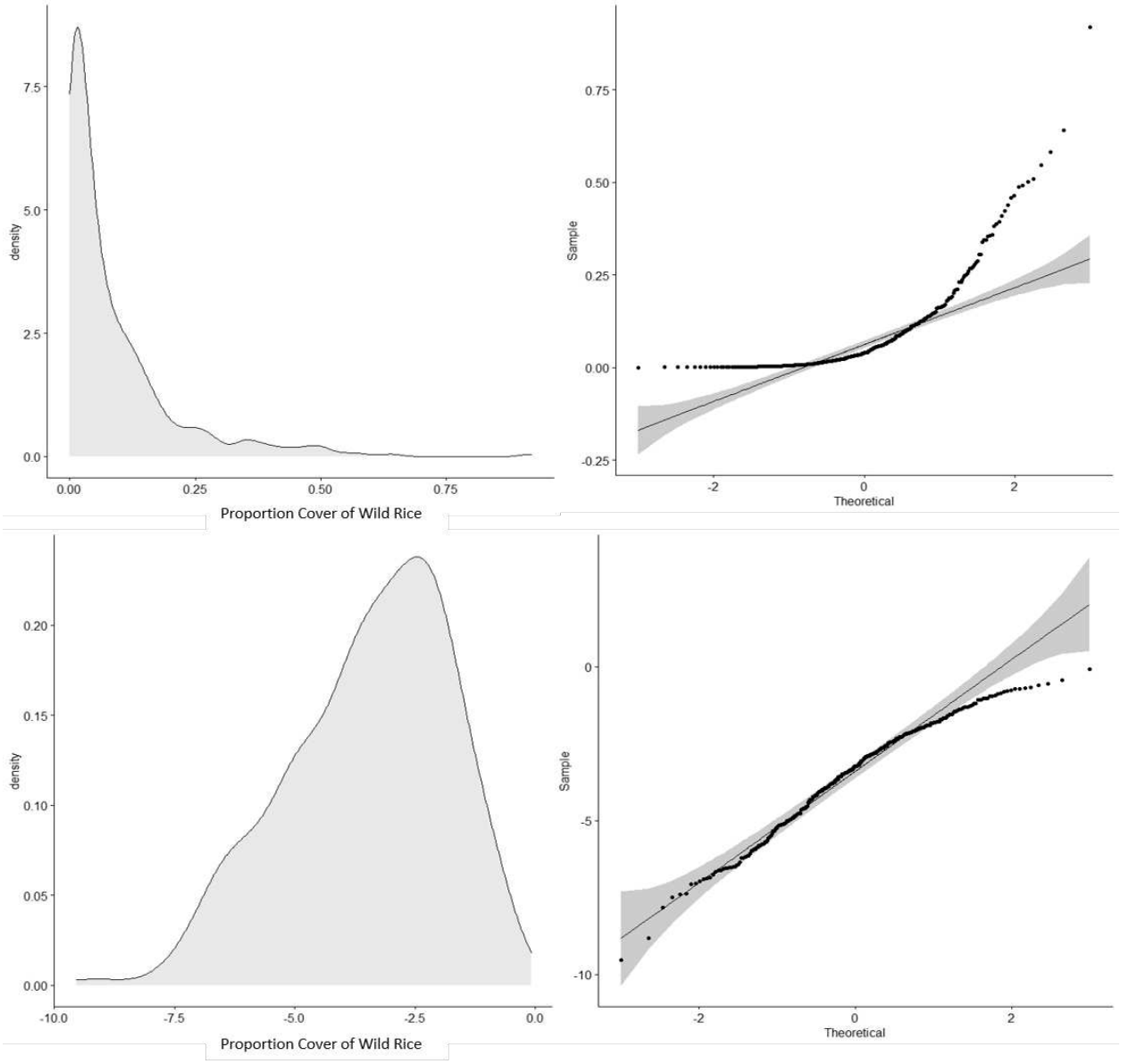


Figure 2.3. Distribution and normality of the proportion cover of wild rice before (top left & right) and after (bottom left & right) log transformation.

Predictor Variables

We utilized a suite of existing spatial datasets to assess wild rice cover as it relates to bathymetry, land cover, topography, and watershed characteristics. Bathymetric data was utilized to derive the average and maximum depth of each lake (Appendix A). These variables may help indicate how susceptible a given lake is to eutrophication at a degree that has significant impacts on water clarity, as well as potentially indicating how susceptible a lake is to dramatic temperature fluctuations, or reaching freezing temperatures long enough to

break seed dormancy of wild rice. The National Agricultural Statistics Service land cover dataset from 2018 was acquired for the study region at a 30 m spatial resolution and reclassified into 11 classes (Appendix A). A single, condensed class represented crops, and all original developed land cover classes were included as potential predictors to provide information about the intensity of urbanization. Each land cover class was summarized by its proportion of cover, and this was done within each catchment and HUC12 watershed, as well as 500, 1,000, and 2,000 m buffers around each lake. Preliminary investigations demonstrated that relationships between land cover and wild rice abundance were strongest when land cover was summarized by catchment regions, which was the scale utilized to investigate land cover in all subsequent analyses.

Topographic gradients often correspond with spatial positions of features within a watershed; therefore, we measured elevation within each lake (Appendix A) using a 30 m digital terrain model constructed from LiDAR and NED 10 m data (Minnesota Department of Transportation, 2017). Catchment, flowline, and lake attributes were extracted from the NHDPlus dataset version 2.1 (U.S. Geological Survey, 2019). Variables included information about water quality (reciprocal hydrologic area load), flow rate, flow volume, total distance of stream paths to the top of the watershed (arbolate sum), distance to terminal flowlines, and length of each flowline (Appendix A). We retrieved detailed watershed health assessment metrics from the Watershed Health Assessment Framework (WHAF) provided by the Minnesota Department of Natural Resources. WHAF captures a variety of metrics summarized by HUC08 watersheds, and many of these metrics were also summarized at the catchment level (Watershed Health Assessment Framework, n.d.). Watershed health assessment scores were ranked from 0 to 100 and incorporated a large suite of datasets related to biology, connectivity, geomorphology, hydrology, and water quality.

We also derived metrics to capture land surface temperature (LST) and the spatial configuration of wild rice lakes. LST was extracted from MODIS Terra daily global emissivity and land surface temperature (MOD11A1 V6) at 1 km spatial resolution using Google Earth Engine to serve as a proxy for water temperature. We created image composites between December 20th - February 20th in 2017 and also June 15th - August 15th 2018 to capture the maximum summer temperature, minimum winter temperature, mean maximum summer temperature, and mean minimum winter temperature (Appendix A).

Dispersal is an important biological mechanism that influences observed patterns of species abundance and distribution, particularly the capacity of hydrologic dispersal for wetland and emergent aquatic vegetation species (Soomers et al., 2012). To account for dispersal influences within our analyses, various metrics were calculated to capture patterns of clustering in wild rice distribution using all lakes with predicted wild rice presence (including all lakes without available bathymetry data). We calculated the euclidean distance of the minimum, average, and maximum distance of the nearest neighbor lake containing wild rice (out of the 10 nearest neighbors) for each sample lake. Through initial spatial explorations of wild rice cover, it tended to be highly clustered which was evident through visual inspection and supported statistically (Moran's $I = 0.29$, $\alpha < 0.05$, p-value < 0.0001). A variable to capture the significance of these hot spots was created to provide more detailed spatial information about the dispersal mechanisms of wild rice. There were 17 lakes that contained over 100 acres of wild rice, and these lakes were classified as the key dispersal hot spots. The euclidean distance to the nearest hot spot lake was also calculated for each of the 366 lakes included in our final analyses. We extracted all predictor values for each sample lake to incorporate into the subsequent modeling workflow.

Modeling Framework Step 1) Variable Selection

Random forest and variable selection algorithms served as tools for preliminary investigation to identify the most important variables, explore correlations between variables, and determine how well the individual predictors described wild rice cover (Figure 2.4). We selected model variables utilizing a balance between algorithmic and human input (Young et al., 2020). This includes ranking variables based on their correlation and explanatory power, while also considering ecological significance from previous literature within variable reduction procedures. A total of 165 variables were distributed into four classes relating to land cover, dispersal, variables measured at the lake or catchment scale, and variables measured at the HUC08 watershed scale. We ranked each class of variables using the rfUtilities package (version 2.1-5; Evans et al., 2011) in R (version 3.6.3; R Core Team, 2020), and incorporated the top four uncorrelated variables (Pearson's correlation coefficient < 0.7) from each class of predictors in the final dataset. We then ranked the

refined dataset consisting of 16 predictors (four variables from each of the four classes) through a final iteration of rfUtilities, and removed any additional variables with a Pearson's correlation coefficient > 0.7. The remaining 10 predictor variables were incorporated into a random forest model to explore their relative predictive power (randomForest package, version 4.6-14) in R, and to incorporate them into the subsequent stages of the modeling framework.

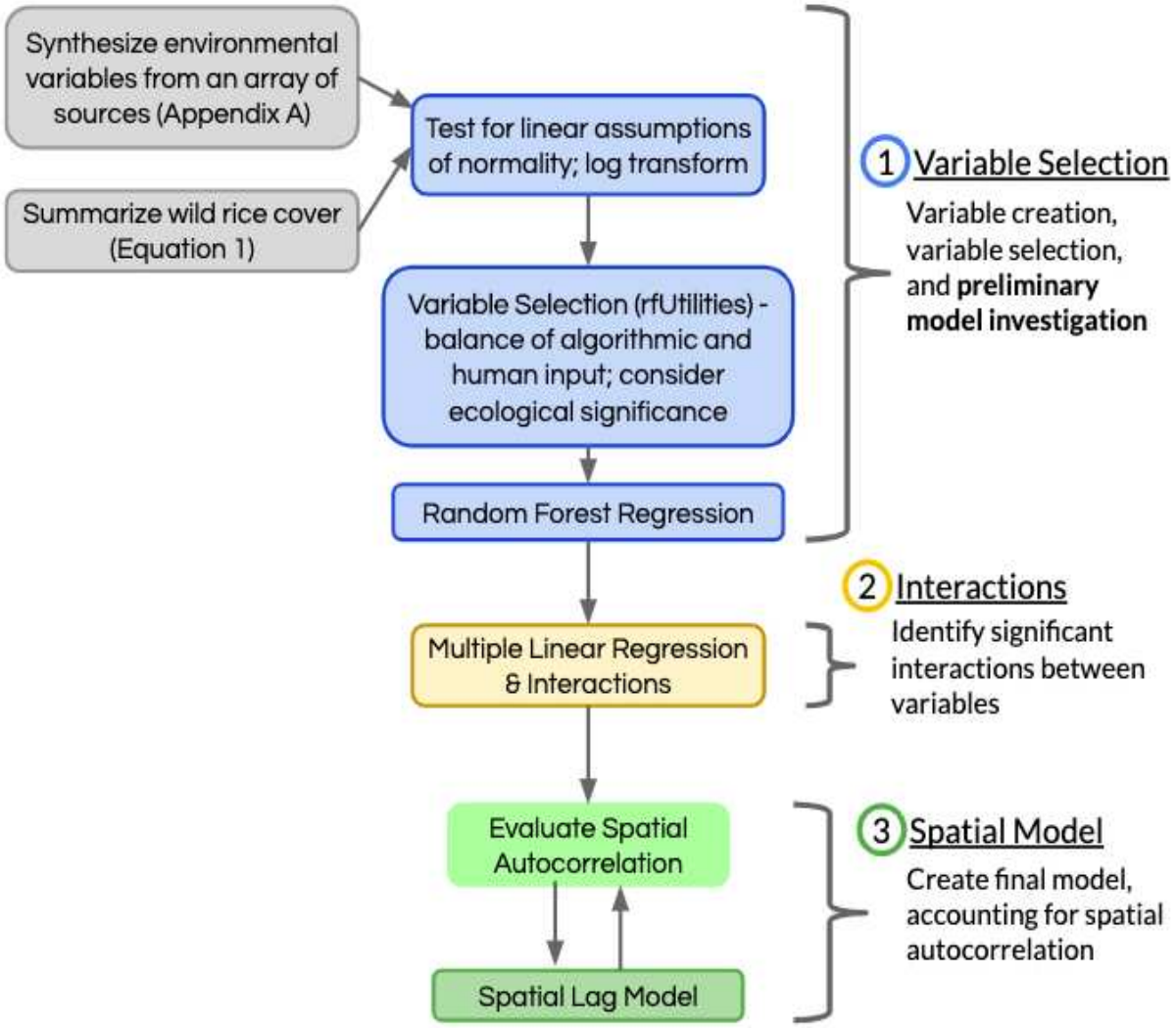


Figure 2.4. The three-step modeling framework utilized in this study which includes: 1) Variable Selection; 2) Interactions; and 3) Spatial Model. Predictors from the random forest model were incorporated into a multiple linear regression to test for statistically significant interactions, and the variables and selected interaction terms were utilized to fit a spatial lag model and account for spatial autocorrelation.

Modeling Framework Step 2) Interactions

Our next objective was to specifically investigate the potential interactions between the predictor variables as well as their directional relationships with one another, to overcome some of the challenges with interpreting interactions within random forest models (Denisko & Hoffman, 2018). We utilized the top variables selected through the random forest model exploration in a multiple linear regression in R to investigate potential interactions between related environmental processes (Figure 2.4; Miller, Franklin, & Aspinall, 2007). We applied a logarithmic transformation to all variables that did not meet the appropriate linear assumptions of normality. Specific interactions were evaluated between related and interdependent environmental variables. For example, we initially tested for relationships between average lake depth and maximum surface temperature since larger and deeper lakes may experience smaller diurnal ranges in temperature (although a significant interaction did not exist). A one-way ANOVA was used to evaluate the inclusion of interaction terms in the multiple linear regression (interaction model) to determine statistical significance in model performance. Our final criteria for the inclusion of an interaction term was a clear ecological relationship between the two interacting variables.

Modeling Framework step 3) Spatial Model

We tested the residuals from the interaction model for spatial autocorrelation using Moran's I (Figure 2.4). The residuals were indicative of spatial autocorrelation and had statistically significant p-values for Moran's I, warranting a spatial modeling framework. Wild rice cover within suitable habitat also demonstrated spatial autocorrelation with a statistically significant Moran's I (0.29, p-value < 0.0001). A spatial lag model was considered for the best interaction model since spatial lag models are known to best capture dispersal, disturbance, or other fine scale spatial dependencies between environmental variables (Miller, Franklin, & Aspinall, 2007; Lichstein et al., 2002). We compared the interaction and spatial lag models based on their Akaike's Criterion Score (AIC) score, where the model with the lowest AIC indicated the best fit. All spatial modeling was completed using the spdep (version 1.1-3; Bivand & Wong, 2018) and spatialreg (version 1.1-5; Bivand, Hauke, & Kossowski, 2013) packages in R.

Results

Random Forest Variable Selection

Our preliminary investigation within the first step of our 3-step modeling framework revealed 10 important environmental variables which explained approximately 36% of the variance within the exploratory random forest model (Figure 2.5). The most important variable was the distance of a given wild rice lake to the nearest wild rice hot spot lake, and this variable alone explained more than 20% of the variance within the model (Table 2.1; Figure 2.5). The water quality local pollution index measured within catchments also contributed substantial explanatory power (Figure 2.5). Geomorphological susceptibility to pollution, average depth of lakes, and flow variability contributed nearly even degrees of explanatory power. These variables were also closely followed by the maximum summer surface temperature, the minimum nearest neighboring wild rice lake, and the percent cover of wetlands in the catchment. The arbolate sum contributed minimally to the model's explanatory power, and the reciprocal area hydraulic load contributed the least to the model.

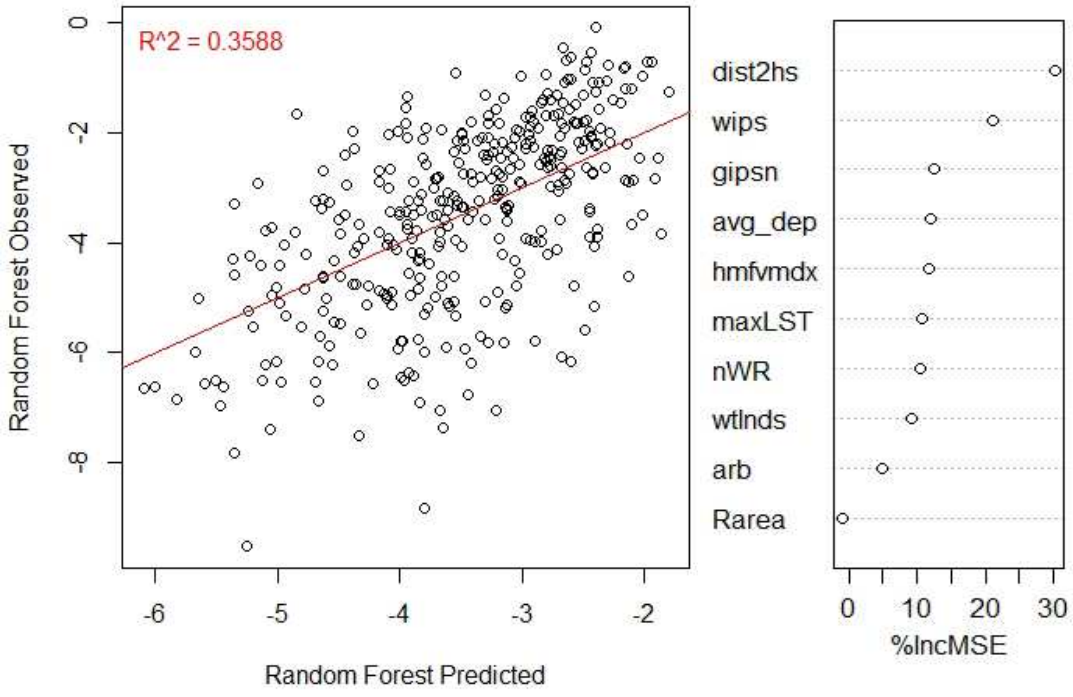


Figure 2.5. Random forest model predicted and observed values (left) with an RMSE of 1.82 and R^2 of 0.359. Ranked variable importance as given by the increase in mean square error (right).

Table 2.1. The final variables selected through the first step within the three-step modeling framework (Figure 2.4).

Variable Abbreviation	Variable Name	Definition	Scale/Units
dist2hs	Distance to Hot Spot	The euclidean distance of a lake to a wild rice hot spot lake, one of 17 lakes in the study region containing 100+ acres of wild rice in 2018.	Lakes, euclidean distance in m
wips	Water Quality Local Pollution Index	Local known pollution sources (superfund sites, potential contaminants, feedlots, open mine pits, and wastewater discharge permits). Index values span 0 to 100, high values represent healthy regions without known pollution sources. Calculated based on the watershed mean value (Watershed Health Assessment Framework, n.d.) https://www.dnr.state.mn.us/whaf/about/scores/water_quality/point.html	Catchment mean score; index
avg_dep	Average Depth	Absolute value of the average depth of a lake	Lake, ft
hmfvmdx	Flow Variability Index	Hydrologic flow variability referencing the annual duration and magnitude of extremes. Lower values indicate more severe alteration to flow or dramatic variability of extremes (Watershed Health Assessment Framework, n.d.). https://www.dnr.state.mn.us/whaf/about/scores/hydrology/flowvariability.html	HUC08 score
gipsn	Geomorphological Pollution Sensitivity Index	Groundwater susceptibility to pollution based on geomorphic setting and the nearest surface materials. It is calculated based on mean watershed values. Lower values indicate greater susceptibility. (Watershed Health Assessment Framework, n.d.) https://www.dnr.state.mn.us/whaf/about/scores/geomorphology/g_i_psnsm.html	Catchment mean score; index
nWR	Minimum Nearest Neighbor	The minimum distance of the nearest wild rice lake	Lakes, euclidean distance in m
wtlnds	Percent Cover of Wetlands	The percent cover of wetlands within the catchment region (wetlands area / catchment area)	Catchment, % cover
maxLST	Maximum Summer Temperature	The maximum summer land surface temperature between June 15th and August 15th of 2018	Lakes, median ºC
arb	Arbolate Sum	Cumulative distance of all upstream paths (U.S. Geological Survey, 2019).	Distance, m
Rarea	Reciprocal Area Hydraulic Load	"Hydraulic metric of time required to displace one unit volume of water", (see Schmadel et al., 2018).	Catchment mean values; ft/day

Multiple Linear Regression & Interactions

All the important variables identified through the random forest exploration were incorporated into a multiple linear interaction model. A significant interaction was identified between the water quality local pollution index and the geomorphological pollution sensitivity index. There was also a significant interaction between the reciprocal area hydraulic load (Schmadel et al., 2018) and the arbolate sum (U.S. Geological Survey, 2019). The linear interaction model had an adjusted R^2 of 0.362, a significant p-value (< 0.0001), and all variables were statistically significant as well (Table 2.2).

Table 2.2. Estimated parameters for all variables included in the interaction model. Note that only covariates that did not meet the assumptions of a linear regression were log transformed, and parameter estimates were not back log transformed to preserve the directional relationships of the main effects with wild rice cover. Red indicates an estimate that is not log transformed and green indicates an estimate that has been log transformed. (p-value < 0.05 = * | p-value <0.01 = ** | p-value < 0.001 = *** | p-value <0.0001 = ****)

Coefficients	Estimate	P value	Log Transform
Intercept	-1.79	****	n/a
Distance to Hotspot	-2.703	****	False
Maximum Surface Temperature	0.3278	****	False
Average Depth	-0.2146	*	True
Percent Cover of Wetlands	1.591	*	False
Reciprocal Area Hydraulic Load	0.2007	**	True
Arbolate Sum	0.1812	***	True
Water Quality Local Pollution Index	0.06928	****	False
Geomorphological Pollution Sensitivity Index	0.09366	***	False
Flow Variability Index	0.16	**	True
Minimum Nearest Neighbor	-0.0005463	****	False
Arbolate Sum * Reciprocal Area Hydraulic Load	-0.1032	*****	True
Water Quality Local Pollution Index * Geomorphological Pollution Sensitivity Index	-0.001132	***	False

Spatial Lag Model

We found statistically significant spatial autocorrelation within the residuals of the interaction model using Moran's I (0.106, p-value =0.009), and therefore continued with fitting a spatial lag model using the same variables and interaction terms. The spatial lag model had an AIC score of 1,266.4 and improved upon the interaction model which had an AIC score of 1,272.7. It adequately addressed spatial autocorrelation within the residuals from the interaction model (Figure 2.6; Figure 2.7), and the residuals of the spatial lag model had a Moran's I value of 0.006 with a statistically insignificant p-value (0.391). The spatial autoregressive parameter for the model was 0.21 and the likelihood ratio of this model was significant (8.3, p-value = 0.004). The distance to the nearest hot spot remained the strongest predictor within this model, however, the spatial lag model revealed a lower parameter estimate than the interaction model, and this occurred with many of the other variables as well (Table 2.2; Table 2.3). This suggests that spatial autocorrelation within the interaction model had an effect on the coefficient estimates.

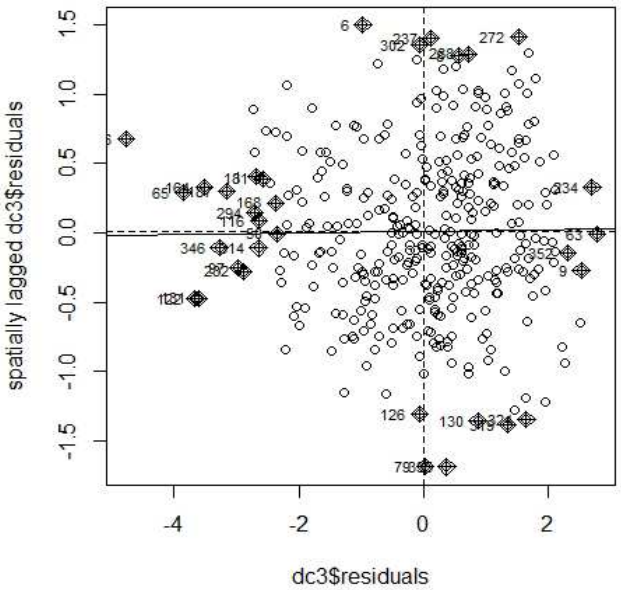


Figure 2.6. Plotted residuals from the spatial lag model. The lack of trend within the residuals exemplifies the mitigation of spatial autocorrelation through the fitting of a spatial lag model.

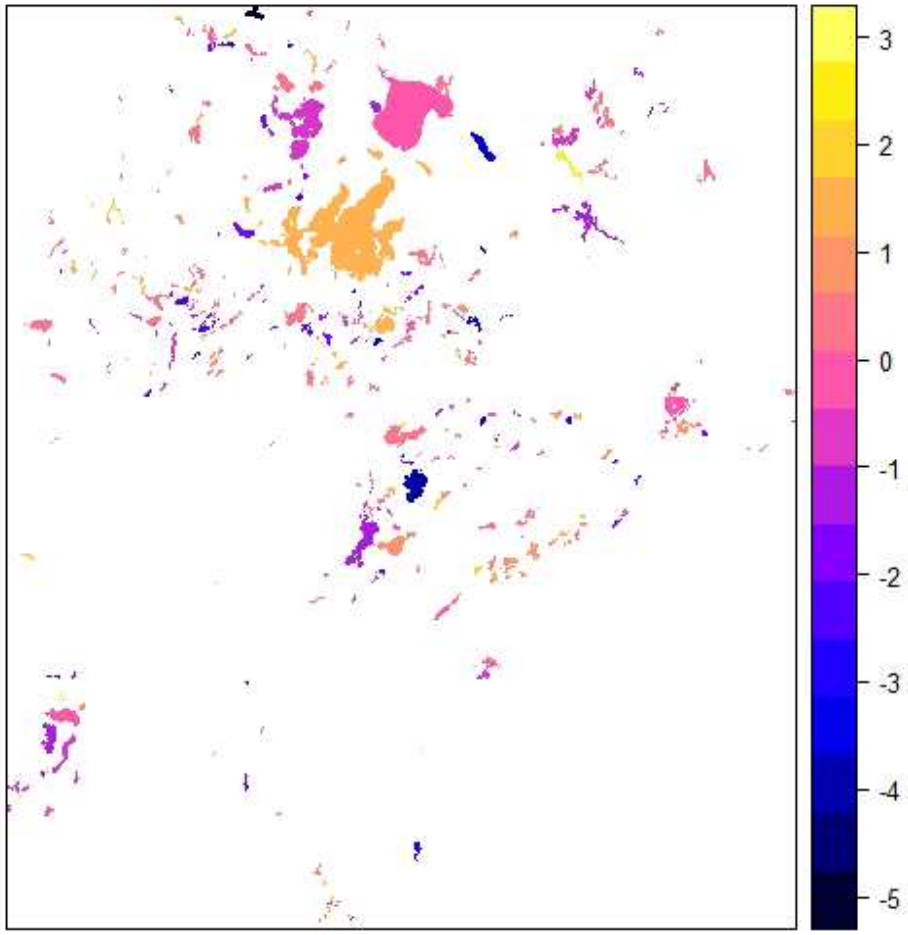


Figure 2.7. Map of residuals from the spatial lag model, where the lack of spatial pattern in residuals further affirms that spatial autocorrelation was adequately addressed by fitting a spatial lag model.

Table 2.3. Estimated parameters for all variables included in the spatial lag model. Red indicates an estimate that is not log transformed and green indicates an estimate that has been log transformed. (p-value < 0.05 = * | p-value <0.01 = ** | p-value < 0.001 = *** | p-value <0.0001 = ****).

Coefficients	Estimate	p-value	Log Transform
Intercept	-15.881	****	n/a
Distance to Hotspot	-2.1271	****	False
Maximum Surface Temperature	0.30097	***	False
Average Depth	-0.20753	**	True
Percent Cover of Wetlands	1.8075	**	False
Reciprocal Area Hydraulic Load	0.20477	***	True
Arbolate Sum	0.185	***	True
Water Quality Local Pollution Index	0.061072	****	False
Geomorphological Pollution Sensitivity Index	0.08087	**	False
Flow Variability Index	0.13753	*	True
Minimum Nearest Neighbor	-0.00052123	***	False
Arbolate Sum * Reciprocal Area Hydraulic Load	-0.10371	*****	True
Water Quality Local Pollution Index * Geomorphological Pollution Sensitivity Index	-0.00097651	***	False

Within the spatial lag model (Table 2.3), the distance to the nearest hot spot exhibited a negative relationship with wild rice cover, demonstrating that wild rice cover decreases with increasing distance to the nearest hotspot, and the minimum nearest neighboring wild rice lake also exhibited this relationship. This relationship was expected, since wild rice cover was highly clustered and had a significant Moran's I value. Increased wild rice cover was associated with higher cover of wetlands in the surrounding landscape and more shallow (average) lake depths, and also exhibited a slightly positive relationship with higher maximum surface temperatures. Wild rice was negatively related to the interaction between the water quality local pollution and geomorphological sensitivity indices, as well as the interaction between reciprocal area hydraulic load and arbolate sum.

Discussion

Through the development of our three-step modeling approach, we identified directional relationships of 10 significant systemic drivers and two interaction terms in relation to wild rice cover and

adequately addressed spatial autocorrelation within our data. The significant drivers generally included lake-specific metrics, characteristics of catchments, characteristics of the HUC08 watersheds, and regional patterns of surface temperature. Dependence between interrelated processes was better captured when key interactions were established, and when metrics represented a multitude of spatial scales, unique to the scale at which a specific process occurs.

The best spatial lag model was achieved through a balance of algorithmic predictor selection and moderation via human input with the lowest AIC value of 1,266.4. The spatial lag model selected through algorithmic input alone (using the top ranking 10 uncorrelated variables) yielded a higher AIC value (1292.5) and worse metrics throughout all exploratory models (Appendix B). These results support organizing variable categories using human input based on ecological knowledge of the system during the algorithmic variable selection process (first stage of the modeling framework) to identify the most important drivers (Young et al., 2020) and create a better fitting model. The relationships we explored were based on the proportion of wild rice in suitable habitat, and suitable habitat may have been overestimated by using bathymetric contour lines <10 ft depth (section 2.2) to compensate for uncertainty within hydrologic and bathymetric boundaries. It should be considered that the minimum mapping unit of 11 adjacent pixels predicted as wild rice was utilized to reduce noise from remotely sensed predictions of presence. This may have misrepresented small or disconnected patches of wild rice, and the significant variables identified may be more strongly correlated with larger, connected patches of wild rice (Fournier et al., 2007).

The key drivers identified within our analyses were generally consistent with previous literature describing environmental influences on wild rice. The distance to the nearest wild rice hotspot was a significant variable that demonstrated wild rice lakes followed patterns of clustering as we had expected (based on the significant Moran's I value). Previous studies have not considered dispersal of wild rice alone, but dispersal was suggested to have an influence on its distribution (DNR, 2008). The strong correlation between wild rice cover and the distance of lakes to the nearest hotspot location may support this observation, and is further supported by the significance of the distance to the nearest neighboring lake containing wild rice. Dispersal is one of the most important mechanisms driving aquatic macrophyte

assemblages and the results may suggest that dispersal limitations may serve as a top down control on the abundance and distribution of wild rice, which would align with results of similar studies conducted for other vegetation species (Robledo-Arnuncio et al., 2014; Mikulyuk et al., 2011; Flinn et al., 2010). The distance to the nearest hot spot metric may also reflect a lake's degree of isolation, population fitness, or cross wind pollination. For instance, Lu, Waller, & David (2005) found that higher connectivity between wild rice populations may improve their resilience. The importance of connectivity between wild rice populations supports cooperative management efforts at a broader scale than monitoring individual lakes; thus, land managers should consider the degree of connectivity between lakes and their spatial configuration on the landscape during conservation planning for the species. It may be more appropriate to view populations as connected habitat regions, in contrast to the more widely accepted notion that wild rice lakes are simply independent and self-contained populations.

Previous studies have recognized that wild rice seeds require multiple consecutive months of freezing temperatures to germinate the following year (Myrbo et al., 2017(a); Kahler et al., 2014; Lee, 2002; Kovach & Bradford, 1992). We found that the maximum summer surface temperature was positively correlated with wild rice cover within our study, which may suggest that wild rice is tolerant of warmer summer temperatures within shallow lakes, and that these shallower lakes would be more likely to reach an appropriate duration of winter freezing temperatures that break seed dormancy. However, MODIS LST had a coarser spatial resolution (1 km) than many of the other important drivers we identified, and land surface temperature was not examined on lakes completely void of wild rice cover; thus these findings should be interpreted with caution as further investigation is needed to untangle the full relationship between wild rice and surface temperature.

Within our study, higher wild rice cover was negatively associated with average lake depth. Similarly, Pillsbury & McGuire (2009) found that lower density wild rice stands tended to occur in deeper water and considered that deeper sites may be prone to more disturbance or additional stress; Meeker (2000) drew similar inferences attributing deeper water with increased disturbance. While additional exploration is needed, another possible explanation is that shallow lakes with a lower average depth may provide larger patches of

suitable habitat that have greater connectivity and more even bathymetric slopes which may allow for greater rates of cross pollination between stands. Deeper lakes with higher average depths may contain steeper bathymetric gradients, smaller patches of suitable habitat, and lower abundance of wild rice; these conditions may be less favorable for wild rice because if the water level rises it may quickly become unsuitable habitat, eliminating smaller, disconnected, and less stable populations. Future research should consider investigating connectivity between patches of suitable habitat, abundance, and connectivity between patches of wild rice to assess stability of populations. Since wild rice is found to prefer lower average lake depths, it is imperative to manage these lakes more carefully for water clarity and quality due to the sensitivity of shallow lakes (Gulati et al., 2007).

Wild rice is favored by brief, low magnitude, and infrequent natural disturbances in hydrologic flows which allow it to outcompete waterlilies (Meeker & Tillison, 2018; Myrbo et al., 2017(a); Pillsbury & McGuire, 2009). However, high magnitude, frequent, and persistent natural disturbances in hydrologic flows can introduce sediment loads that interfere with wild rice seed germination. Alternatively, disturbances during the vulnerable floating leaf stage can negatively impact wild rice (May to mid June; Myrbo et al., 2017(a); DNR, 2008). Anthropogenic manipulation of hydrologic systems alters flow rates and leads to sporadic, poorly timed, extreme variation in flows, and larger sediment loads that tend to negatively impact wild rice (DNR, 2008). Therefore, we expected that wild rice would be found nearest to headwaters based on cumulative effects of sediment and nutrient loads downstream, and thus wild rice cover would exhibit a negative relationship with arbolate sum. We also hypothesized that wild rice cover would follow a Gaussian distribution in response to reciprocal area hydraulic load; implying that wild rice would have a positive relationship with reciprocal area hydraulic load up until a certain flow rate threshold, and would have a negative relationship with reciprocal area hydraulic load when water is transferred faster than a certain rate. Less extreme flow variability, or unaltered flows, were represented by higher values in the flow variability index, and we thought this index would be positively associated with wild rice cover. In support of our initial expectations, the flow variability index was positively correlated with wild rice cover since dams, ditches, and other manipulations of hydrologic flows negatively impact wild rice. Additionally, the interaction between

daily water displacement (reciprocal area hydraulic load) and cumulative stream path distance from headwaters (arbolate sum) had a statistically significant negative effect on wild rice cover and supported our initial expectations. The significant negative interaction effect may demonstrate that the true relationship with wild rice cover stem from the interdependence between these two related factors. It is possible that the accumulation of negative impacts downstream are not captured by the individual metrics. These metrics all demonstrate that wild rice is negatively impacted by altered hydrologic flows, and the impacts have greater negative effects downstream. This may be an important relationship to consider in the larger scheme of hydrologic management. A connected and holistic management plan across the landscape may optimize efforts to preserve flow rates that benefit wild rice, with consideration of its distribution and habitat preferences.

We found two significant interaction terms (water quality local pollution and the geomorphologic pollution sensitivity indices; arbolate sum and reciprocal area hydraulic load) that may also suggest that there is not only one specific pollution type or topographical predisposition that influences the presence and abundance of wild rice on a given lake. This indicates that wild rice is more likely dependent on multiple interrelated factors, and their specific degrees of negative or positive influence corresponds to more than a single hydrologic or geomorphic metric (Myrbo et al., 2017(a); DNR, 2008). It further reinforces that it is imperative to address the intimate relationships between anthropogenic influences and the innate complexity of environmental drivers from a holistic perspective that considers the interactions between landscape drivers as well as the relative importance of space and scale.

Previous studies have recognized that dispersal plays a critical role predicting the adaptation of vegetation to changing climate (Robledo-Arnuncio et al., 2014; Mikulyuk et al., 2011). Considering that dispersal or population connectivity had the strongest relationship with wild rice cover, it may be worth further investigating these traits further to understand how wild rice may be impacted by changing climate. Wild rice is constrained by its shattering traits and primarily dispersed by waterfowl, wind (with variable direction), or hydrologic flows generally flowing toward the south; these mechanisms may present challenges for populations to migrate northward and remain with suitable climate envelopes (Loarie et al., 2009). Wild

rice positioned at the headwaters may also face challenges migrating into Northbound watersheds due to less suitable water quality at the terminal ends of the watershed. Essentially, it is unclear how wind and waterfowl will disperse wild rice, but within the study region, hydrologic flows will primarily distribute wild rice south (and slightly east) of the current hotspots, presenting potential migration barriers. Future studies should explore the relative influences of wind and waterfowl on wild rice dispersal and the implications of adaptability or dispersal in response to changing climates.

Conclusion

This study employed a three-step modeling framework to select the most influential drivers, identify interactions between drivers, and account for spatial patterns in relation to wild rice cover across lakes in Minnesota. We found that wild rice abundance and distribution is primarily a result of fine-scale systematic drivers like dispersal or population connectivity, and coarse-scale drivers of temperature, flow regimes, and other anthropogenic influences play significant, but less important roles. It may be essential for landscape managers to consider wild rice lakes from a matrix perspective of connected habitats and populations in contrast to managing and monitoring individual lakes as if they are independent of one another. The mechanisms and patterns of dispersal may be crucial for predicting future distributions of wild rice in response to changes in climate or identifying suitable habitat suitable under future conditions. Factors influencing dispersal limitations should be investigated in greater detail to identify potential mechanisms or patterns that were not captured by the datasets utilized in this study. This may include the effects of wind pollination, size of suitable habitat patches and specifically, the connectivity between patches in relation to abundance patterns through time, as larger populations tend to exhibit greater trait diversity and resilience (Lu, Waller, & David, 2005). Furthermore, future studies should consider expanding these analyses to broader regional extents to examine the influences of systemic drivers to determine the scales at which each is most relevant, or perhaps if patterns of additional regional top-down drivers were not identified at the spatial extent we utilized. Multiyear studies analyzing the impacts of changes in annual precipitation, as well as

surface temperature extremes and durations may reveal how wild rice will respond to changing climate to establish proactive conservation measures. In addition to the known relationship between wild rice and pore water sulfide/surface water sulfate, subsequent investigations have the potential to more explicitly quantify the influence of other primary controls on wild rice abundance relevant to conservation and restoration planning.

REFERENCES

Bivand, R. S., Hauke, J., and Kossowski, T. (2013). Computing the Jacobian in Gaussian spatial autoregressive models: An illustrated comparison of available methods. Geographical Analysis, 45(2), 150-179. URL https://doi.org/10.1111/gean.12008

Bivand, Roger S. and Wong, David W. S. (2018) Comparing implementations of global and local indicators of spatial association TEST, 27(3), 716-748. URL https://doi.org/10.1007/s11749-018-0599-x

Blann, Kristen L., Anderson, James L., Sands, Gary R., & Vondracek, Bruce (2009). Effects of Agricultural Drainage on Aquatic Ecosystems: A Review, Critical Reviews in Environmental Science and Technology, 39:11, 909-1001, DOI: 10.1080/10643380801977966

Cressie, N.A.C. (1993) Statistics for spatial data. Wiley, New York.

Delouche, J. C., & Bugros, N. R. (2007). Weedy rices-origin, biology, ecology and control. Rome: FAO.

Denisko, D., & Hoffman, M. M. (2018). Classification and interaction in random forests. Proceedings of the National Academy of Sciences, 115(8), 1690-1692. doi:10.1073/pnas.1800256115

Dormann, C. F. (2007). Effects of incorporating spatial autocorrelation into the analysis of species distribution data. Global Ecology and Biogeography, 16(2), 129–138. doi: 10.1111/j.1466-8238.2006.00279.x

DNR (2008). Natural wild rice in Minnesota, Minnesota Department of Natural Resources. St. Paul, Minnesota. Retrieved from http://files.dnr.state.mn.us/fish_wildlife/wildlife/shallowlakes/natural-wild-rice-in-minnesota.pdf (accessed August, 2018)

Evans JS, Murphy MA, Holden ZA, & Cushman SA (2011). "Modeling species distribution and change using Random Forests." In Drew CA, Wiersma YF, Huettmann F (eds.), _Predictive species and habitat modeling in landscape ecology: concepts and applications_, chapter 8, 139-159. Springer, New York. ISBN 978-1-4419-7390-0.

Fischer, E. M., & Knutti, R. (2015). Anthropogenic contribution to global occurrence of heavy-precipitation and high-temperature extremes. Nature Climate Change, 5(6), 560–564. doi: 10.1038/nclimate2617

Flinn, K. M., Gouhier, T. C., Lechowicz, M. J., & Waterway, M. J. (2010). The role of dispersal in shaping plant community composition of wetlands within an old-growth forest. Journal of Ecology, 98(6), 1292–1299. doi: 10.1111/j.1365-2745.2010.01708.x

Forman, R. T. T., & Gutzwiller, K. J. (2002). Applying landscape ecology in biological conservation. New York: Springer.

Fournier, R. A., Grenier, M., Lavoie, A., & Hélie, R. (2007). Towards a strategy to implement the Canadian Wetland Inventory using satellite remote sensing. Canadian Journal of Remote Sensing, 33(sup1). doi: 10.5589/m07-051

Gripenberg, S., & Roslin, T. (2007). Up or down in space? Uniting the bottom-up versus top-down paradigm and spatial ecology. Oikos, 116(2), 181–188. doi: 10.1111/j.0030-1299.2007.15266.x

Gulati, R. D., Lamens, E., De Pauw, N., & Van Donk, E. (2007). Shallow lakes in a changing world: proceedings of the 5th International Symposium on Shallow Lakes, held at Dalfsen, the Netherlands, 5-9 June 2005. Dordrecht: Springer.

Kahler A., Kern A., Porter R., Phillips R. (2014) Maintaining Food Value of Wild Rice (Zizania palustris L.) Using Comparative Genomics. In: Tuberosa R., Graner A., Frison E. (eds) Genomics of Plant Genetic Resources. Springer, Dordrecht

Kennard, W., Phillips, R., Porter, R., Grombacher, A., & Phillips, R. L. (1999). A comparative map of wild rice (Zizania palustris L. 2n=2 x =30). TAG Theoretical and Applied Genetics, 99(5), 793-799. doi:10.1007/s001220051298

Kennard, W., Phillips, R., & Porter, R. (2002). Genetic dissection of seed shattering, agronomic, and color traits in American wildrice (Zizania palustris var. interior L.) with a comparative map. Theoretical and Applied Genetics, 105(6), 1075-1086. doi:10.1007/s00122-002-0988-z

Khoury, C. K., Greene, S., Wiersema, J., Maxted, N., Jarvis, A., & Struik, P. C. (2013). An inventory of crop wild relatives of the United States. Crop Science, 53, 1496–1508.

Kovach, D. A., & Bradford, K. J. (1992). Imbibitional Damage and Desiccation Tolerance of Wild Rice (Zizania palustris) Seeds. Journal of Experimental Botany, 43(6), 747–757. doi: 10.1093/jxb/43.6.747

Lamers, L. P. M., Falla, S.-J., Samborska, E. M., Dulken, I. A. R. V., Hengstum, G. V., & Roelofs, J. G. M. (2002). Factors controlling the extent of eutrophication and toxicity in sulfate-polluted freshwater wetlands. Limnology and Oceanography, 47(2), 585–593. doi: 10.4319/lo.2002.47.2.0585

Lee, P. F. (2002). Ecological relationships of wild rice, Zizania spp. 10. Effects of sediment and among-population variations on plant density in Zizania palustris. Canadian Journal of Botany, 80(12), 1283–1294. doi: 10.1139/b02-118

Levin, S. A. (1992). The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award Lecture. Ecology, 73(6), 1943-1967. doi:10.2307/1941447

Lichstein, J. W., Simons, T. R., Shriner, S. A., & Franzreb, K. E. (2002). Spatial Autocorrelation And Autoregressive Models In Ecology. Ecological Monographs, 72(3), 445–463. doi: 10.1890/0012-9615(2002)072[0445:saaami]2.0.co;2

Loarie, S. R., Duffy, P. B., Hamilton, H., Asner, G. P., Field, C. B., & Ackerly, D. D. (2009). The velocity of climate change. Nature, 462(7276), 1052–1055. doi: 10.1038/nature08649

Lu, Y., Waller, D. M., & David, P. (2005). Genetic variability is correlated with population size and reproduction in American wild-rice (Zizania palustris var. palustris, Poaceae) populations. American Journal of Botany, 92(6), 990–997. doi: 10.3732/ajb.92.6.990

McAtee, W. L. (1917). Propagation of wild duck foods. U. S. Department of Agriculture Bulletin 465. Government Printing Office, Washington, D.C.

Meeker, J. E. 2000. Ecology of "wild" wild rice (Zizania palustris var. palustris) in the Sakagon Sloughs, a riverine wetland on Lake Superior. p. 68–84. In L. S. Williamson, L. A. Dlutkowski, and A. P. McCommon Soltis (eds.) Wild rice research and management. Great Lakes Fish and Wildlife Commission publication, Odanah, WI, USA.

Meeker, J., & Tillison, N. (2018). Kakagon (Bad River Sloughs), Wisconsin (USA). The Wetland Book, 427-435. doi:10.1007/978-94-007-4001-3_229

Mikulyuk, A., Sharma, S., Van Egeren, S., Erdmann, E., Nault, M. E., & Hauxwell, J. (2011). The relative role of environmental, spatial, and land-use patterns in explaining aquatic macrophyte community composition. Canadian Journal of Fisheries and Aquatic Sciences, 68(10), 1778–1789. doi: 10.1139/f2011-095

Miller, J., Franklin, J., & Aspinall, R. (2007). Incorporating spatial dependence in predictive vegetation models. Ecological Modelling, 202(3-4), 225–242. doi: 10.1016/j.ecolmodel.2006.12.012

MNDR.(n.d) Minnesota Climate Trends 1895-2018. Retrieved from https://arcgis.dnr.state.mn.us/ewr/climatetrends/

Minnesota Department of Natural Resources (2014). Lake bathymetric contours. Retrieved from https://gisdata.mn.gov/dataset/water-lake-bathymetry

Minnesota Department of Transportation (2017). Digital Terrain Model - Pits Removed, Minnesota. Retrieved from https://gisdata.mn.gov/lt/dataset/25b6b686-1c25-48e2-8623-62a70ca27823

Muthukrishnan, R., & Larkin, D. J. (2020). Invasive species and biotic homogenization in temperate aquatic plant communities. Global Ecology and Biogeography, 29(4), 656–667. doi: 10.1111/geb.13053

Myrbo, A., Swain, E. B., Engstrom, D. R., Wasik, J. C., Brenner, J., Shore, M. D., Peters, E. P., Blaha, G. (2017) (a). Sulfide Generated by Sulfate Reduction is a Primary Controller of the Occurrence of Wild Rice (Zizania palustris) in Shallow Aquatic Ecosystems. Journal of Geophysical Research:

Biogeosciences, 122(11), 2736-2753. doi:10.1002/2017jg003787

Myrbo, A., Swain, E. B., Johnson, N., Pastor, J., Dewey, B., Engstrom, D. R., ... Peters, E. B. (2017) (b). Increase in nutrients, mercury, and methylmercury as a consequence of elevated sulfate reduction to sulfide in experimental wetland mesocosms. Journal of Geophysical Research: Biogeosciences, 122. https://doi.org/10.1002/2017JG003788

Pastor, J., Dewey, B., Johnson, N. W., Swain, E. B., Monson, P., Peters, E. B., & Myrbo, A. (2017). Effects of sulfate and sulfide on the life cycle of Zizania palustris in hydroponic and mesocosm experiments. Ecological Applications, 27, 321–336.

Pollman, C. D., Swain, E. B., Bael, D., Myrbo, A., Monson, P., & Dykhuizen Shore, M. (2017). The evolution of sulfide in shallow aquatic ecosystem sediments—An analysis of the roles of sulfate, organic carbon, iron and feedback constraints using structural equation modeling. Journal of Geophysical Research: Biogeosciences, 122. https://doi.org/10.1002/2017JG003785

Price, M. W. (2012). Spectral identification of wild rice (*Zizania palustris* L.) using indigenous knowledge and Landsat multispectral data (Master's Thesis). Retrieved from the ScholarWorks at University of Montana Graduate Student Theses, Dissertations, & Professional Papers database. https://scholarworks.umt.edu/etd/910

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.URL http://www.R-project.org/.

Robledo-Arnuncio, J. J., Klein, E. K., Muller-Landau, H. C., & Santamaría, L. (2014). Space, time and complexity in plant dispersal ecology. Movement Ecology, 2(1). doi: 10.1186/s40462-014-0016-3

Scheffer, M. (1998). Ecology of Shallow Lakes. London: Chapman & Hall.

Schmadel, N. M., Harvey, J. W., Alexander, R. B., Schwarz, G. E., Moore, R. B., Eng, K., Gomez-Velez, J. D., Boyer, E. W., & Scott, D. (2018). Thresholds of lake and reservoir connectivity in river networks control nitrogen removal. Nature Communications, 9(1). doi: 10.1038/s41467-018-05156-x

Soomers, H., Karssenberg, D., Soons, M. B., Verweij, P. A., Verhoeven, J. T. A., & Wassen, M. J. (2012). Wind and Water Dispersal of Wetland Plants Across Fragmented Landscapes. Ecosystems, 16(3), 434–451. doi: 10.1007/s10021-012-9619-y

U.S. Geological Survey (2019). National Hydrography Dataset (ver. 2.1 USGS National Hydrography Dataset Plus Attributes for Hydrologic Unit (HU) 7 - 20011), accessed April 6, 2020 at URL https://www.epa.gov/waterdata/nhdplus-upper-mississippi-data-vector-processing-unit-07

Vittori, V. D., Gioia, T., Rodriguez, M., Bellucci, E., Bitocchi, E., Nanni, L., Attene, G., Rau, D., & Papa, R. (2019). Convergent Evolution of the Seed Shattering Trait. Genes, 10(1), 68. doi:10.3390/genes10010068

Watershed Health Assessment Framework (n.d.) Minnesota Department of Natural Resources. Water quality local pollution, geomorphological pollution sensitivity, and flow variability. Retrieved from https://gisdata.mn.gov/dataset/whaf

Young NE, Jarnevich CS, Sofaer HR, Pearse I, Sullivan J, Engelstad P, et al. (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

APPENDICES

Appendix 1A. The number of training samples generated for each EAV class from 2017 and 2018, and the total number of EAV training samples randomly selected for training and validation of the species model.

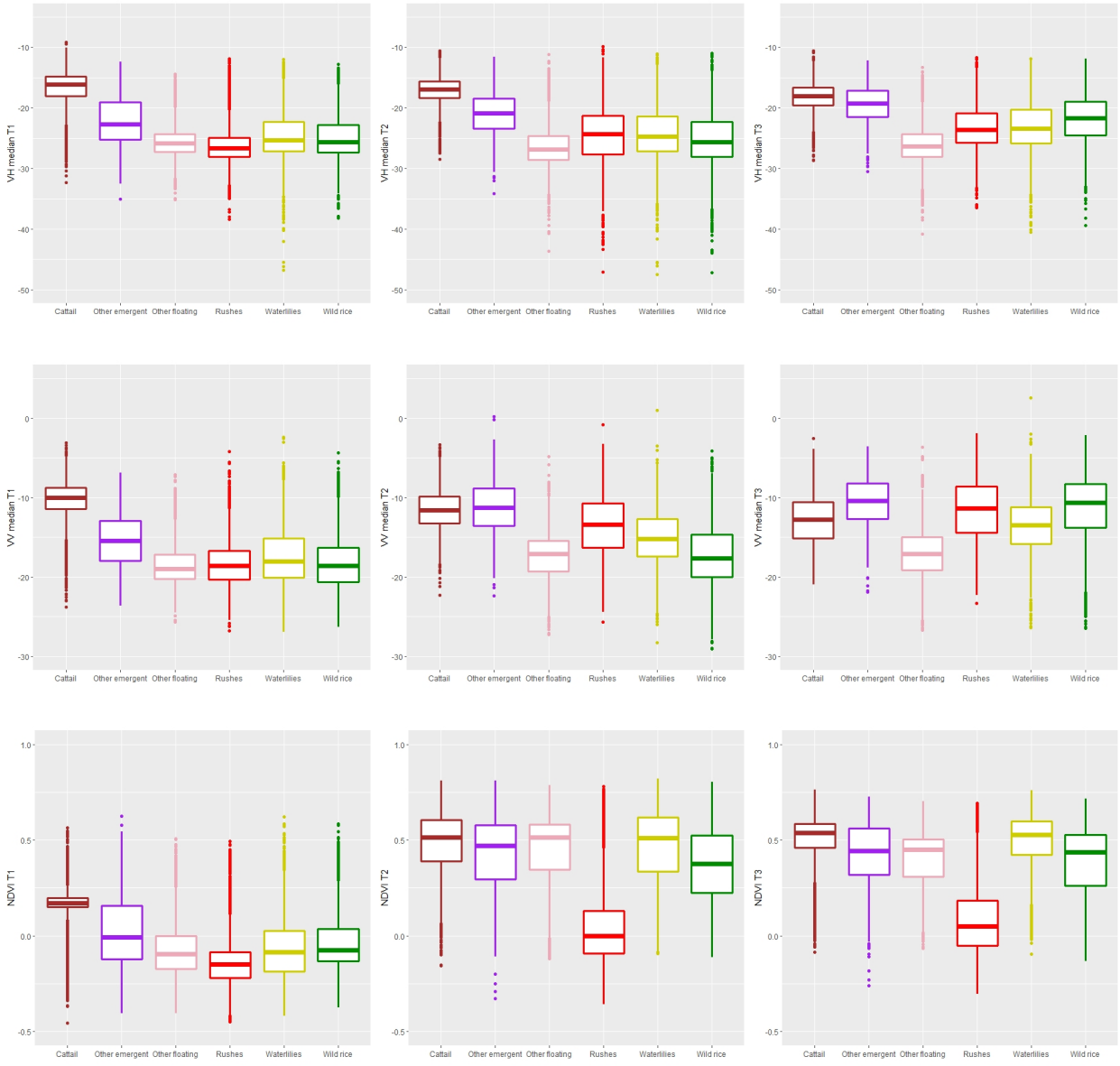
Class	2017 Samples (≥60 m²)	2018 Samples (≥ 60 m²)	Total Number of Samples (≥ 60 m²)	# of Training Samples	# of Validation Samples
Cattails	22,504	16,845	39,349	4,247	1,753
Other EAV	8,468	6,386	14,854	4,241	1,759
Rushes	11,479	50,191	61,670	4,245	1,755
Water lilies	21,894	32,283	54,117	4,245	1,755
Wild rice	2,982	6,555	9,537	4,241	1,759
Total	67,327	112,200	179,527	21,219	8,781

Appendix 1B. Multispectral and SAR indices derived from Sentinel-1 SAR and Sentinel-2 MSI satellite imagery and associated calculations.

Spectral Predictor	Sensor	Spatial Resolution	Calculation
NDVI - Normalized Difference Vegetation Index	Sentinel-2	10 meters	(NIR - Red) / (NIR + Red)
Red Edge #1 NDVI - Normalized Difference Vegetation Index	Sentinel-2	20 meters	(NIR -B5) / (NIR + B5)
Red Edge #2 NDVI - Normalized Difference Vegetation Index	Sentinel-2	20 meters	(NIR -B6) / (NIR + B6)
Red Edge #3 NDVI - Normalized Difference Vegetation Index	Sentinel-2	20 meters	(NIR -B7) / (NIR + B7)
Red Edge #4 NDVI - Normalized Difference Vegetation Index	Sentinel-2	20 meters	(NIR -B8A) / (NIR + B8A)
NDWI - Normalized Difference Water Index	Sentinel-2	20 meters	(Green - SWIR 1) / (Green + SWIR 1)
MNDWI - Modified Normalized Difference Water Index	Sentinel-2	20 meters	(Green - NIR) / (Green + NIR)
B1 Coastal T1 - T2	Sentinel-2	60 meters	B1 median T1 - B1 median T2
B1 Coastal T1 - T3	Sentinel-2	60 meters	B1 median T1 - B1 median T3
B1 Coastal T1 - T4	Sentinel-2	60 meters	B1 median T1 - B1 median T4
B1 Coastal T2 - T3	Sentinel-2	60 meters	B1 median T2 - B1 median T3
B1 Coastal T2 - T4	Sentinel-2	60 meters	B1 median T2 - B1 median T4

B1 Coastal T3 - T4	Sentinel-2	60 meters	B1 median T3 - B1 median T4
B2 Blue T1 - T2	Sentinel-2	10 meters	B2 median T1 - B2 median T2
B2 Blue T1 - T3	Sentinel-2	10 meters	B2 median T1 - B2 median T3
B2 Blue T1 - T4	Sentinel-2	10 meters	B2 median T1 - B2 median T4
B2 Blue T2 - T3	Sentinel-2	10 meters	B2 median T2 - B2 median T3
B2 Blue T2 - T4	Sentinel-2	10 meters	B2 median T2 - B2 median T4
B2 Blue T3 - T4	Sentinel-2	10 meters	B2 median T3 - B2 median T4
B9 Water Vapor T1 - T2	Sentinel-2	60 meters	B9 median T1 - B9 median T2
B9 Water Vapor T1 - T3	Sentinel-2	60 meters	B9 median T1 - B9 median T3
B9 Water Vapor T1 - T4	Sentinel-2	60 meters	B9 median T1 - B9 median T4
B9 Water Vapor T2 - T3	Sentinel-2	60 meters	B9 median T2 - B9 median T3
B9 Water Vapor T2 - T4	Sentinel-2	60 meters	B9 median T2 - B9 median T4
B9 Water Vapor T3 - T4	Sentinel-2	60 meters	B9 median T3 - B9 median T4
VV median	Sentinel-1	10 meters	median
VH median	Sentinel-1	10 meters	median
VV range T1 - T2	Sentinel-1	10 meters	VV median T1 - VV median T2
VV range T1 - T3	Sentinel-1	10 meters	VV median T1 - VV median T3
VV range T1 - T4	Sentinel-1	10 meters	VV median T1 - VV median T4
VV range T2 - T3	Sentinel-1	10 meters	VV median T2 - VV median T3
VV range T2 - T4	Sentinel-1	10 meters	VV median T2 - VV median T4
VV range T3 - T4	Sentinel-1	10 meters	VV median T3 - VV median T4
VH range T1 - T2	Sentinel-1	10 meters	VH median T1 - VH median T2
VH range T1 - T3	Sentinel-1	10 meters	VH median T1 - VH median T3
VH range T1 - T4	Sentinel-1	10 meters	VH median T1 - VH median T4
VH range T2 - T3	Sentinel-1	10 meters	VH median T2 - VH median T3
VH range T2 - T4	Sentinel-1	10 meters	VH median T2 - VH median T4
VH range T3 - T4	Sentinel-1	10 meters	VH median T3 - VH median T4

Appendix 1C. The distribution of predictor values of each class of EAV for VH median (top row), VV median (middle row), and NDVI (bottom row) at T1 (left column), T2 (middle column), and T3 (right column). (Brown = cattails | purple = other emergent; pink = other floating | red = rushes | yellow = water lilies | green = wild rice).



Appendix 1D. The top predictors (T3 only) utilized for the water model ranked by their mean decrease accuracy.

Importance Rank	Mean Decrease Accuracy	Predictor	T1	T2	T3	T4
1	602.285	VV median			X	
2	246.197	NDVI			X	
3	166.623	Blue			X	
4	130.510	NDVI RE #4			X	
5	119.027	Water Vapor			X	

Appendix 2A. All potential predictor variables considered for describing the proportion of cover of wild rice cover within suitable habitat on a given lake in 2018

Predictor	Original Data Source	Description	Scale
Average Depth	Bathymetry	Average lake depth	Lake
Maximum Depth	Bathymetry	Maximum lake depth	Lake
Number of Bathymetry Lines	Bathymetry	The total number of bathymetry lines to highlight the variance in lake depth.	Lake
Number of Flowlines	Bathymetry	The number flowline points entering a lake to highlight the potential number of pollution sources, variation or magnitude of annual flow rates, etc.	Lake
Minimum LST	MODIS Terra 1 km emissivity & LST	The minimum winter land surface temperature.	Lake
Maximum LST	MODIS Terra 1 km emissivity & LST	The maximum winter land surface temperature.	Lake
Mean Minimum LST	MODIS Terra 1 km emissivity & LST	The mean minimum winter land surface temperature.	Lake
Mean Maximum LST	MODIS Terra 1 km emissivity & LST	The mean maximum winter land surface temperature.	Lake
Mean Elevation	Digital Terrain Model - Pits Removed 30 m	Mean elevation of the lake	Lake
Water (% cover)	National Agricultural Statistics Service (NASS) 30m	Area / catchment area	catchment
Fallow Fields (% cover)	National Agricultural Statistics Service (NASS) 30m	Area / catchment area	catchment
Crops (% cover)	National Agricultural Statistics Service (NASS) 30m	Area / catchment area	catchment
Barren (% cover)	National Agricultural Statistics Service (NASS) 30m	Area / catchment area	catchment
Open Developed (% cover)	National Agricultural Statistics Service (NASS) 30m	Area / catchment area	catchment

Low Intensity Developed (% cover)	National Agricultural Statistics Service (NASS) 30m	Area / catchment area	catchment
Medium Intensity Developed (% cover)	National Agricultural Statistics Service (NASS) 30m	Area / catchment area	catchment
High Intensity Developed (% cover)	National Agricultural Statistics Service (NASS) 30m	Area / catchment area	catchment
Grassland/Pasture (% cover)	National Agricultural Statistics Service (NASS) 30m	Area / catchment area	catchment
Shrubland (% cover)	National Agricultural Statistics Service (NASS) 30m	Area / catchment area	catchment
Wetlands (% cover)	National Agricultural Statistics Service (NASS) 30m	Area / catchment area	catchment
Forest (% cover)	National Agricultural Statistics Service (NASS) 30m	Area / catchment area	catchment
Majority Land Cover Type	National Agricultural Statistics Service (NASS) 30m	The land cover class making up the largest area from 2018.	HUC12
Perennial Cover Index	Watershed Health Assessment Framework	Score for perennial vegetation cover in 2001, 2006, and 2011, ranked 0 to 100	catchment, HUC08
Impervious Surface Cover Index	Watershed Health Assessment Framework	Score for impervious surface cover in 2001, 2006, and 2011, ranked 0 to 100	catchment, HUC08
Hydrologic Storage Loss Indices	Watershed Health Assessment Framework	Ratio of the area of unaltered water/streams to the area of altered water/streams; permitted volume of water withdrawal to available runoff/surface water; wetland loss; ratio of average altered water area to wetland loss index value. All indices ranked 0 to 100.	catchment, HUC08
Soil Erosion Susceptibility Index	Watershed Health Assessment Framework	Sediment erosion from environmental processes based on upland areas, channel sediment, and stream banks.	catchment, HUC08
Geomorphological Pollution Sensitivity Index	Watershed Health Assessment Framework	Groundwater susceptibility to pollution based on geomorphic setting and the nearest surface materials. It is calculated based on mean watershed values. Lower values indicate greater susceptibility.	catchment, HUC08
Terrestrial Habitat Quality Index	Watershed Health Assessment Framework	The biological integrity and biodiversity of surrounding terrestrial species (avian, vegetation, etc.) as well as habitat size and shape.	catchment, HUC08
Aquatic Species Quality Index	Watershed Health Assessment Framework	The biodiversity and total number of aquatic species (fish, macroinvertebrates, mussels, etc.)	catchment, HUC08
Riparian Connectivity Index	Watershed Health Assessment Framework	The connectivity, size, and integrity of riparian habitats, corridor sizes and area of undeveloped land.	catchment, HUC08
Aquatic Connectivity Index	Watershed Health Assessment Framework	The connectivity of aquatic habitats, which are negatively impacted by dams, culverts, and bridges which alter sediment inputs and change hydrologic flows.	catchment, HUC08
Water Quality Local Pollution Sources Index	Watershed Health Assessment Framework	Local known pollution sources (superfund sites, potential contaminants, feedlots, open mine pits, and wastewater discharge permits). Index values span 0 to 100, high values represent healthy regions without known pollution sources. Calculated based on the watershed mean value.	catchment, HUC08
Hydrologic Flow Variability Index	Watershed Health Assessment Framework	There were 5 indexes within this category, each measuring one of the following: magnitude by month, change	HUC08

		rate/frequency, magnitude and duration of annual extremes, timing of annual extremes, and the frequency and duration of high and low pulses. All indices ranked 0 to 100.	
Climate Water Balance Index	Watershed Health Assessment Framework	Geomorphological index based on precipitation and evapotranspiration means over 30 years. Index ranked 0 to 100.	HUC08
Distance to Headwaters	Near Analysis; ArcGIS	Euclidean distance to Itasca Lake, the headwaters.	Full ROI
Direction to Headwaters	Near Analysis; ArcGIS	Direction to Itasca Lake, the headwaters.	Full ROI
Distance to Hot Spot	Near Analysis; ArcGIS	Euclidean distance to the nearest of 17 wild rice hot spot lakes with 100+ acres of predicted wild rice in 2018	Full ROI
Direction to Hot Spot	Near Analysis; ArcGIS	Direction to the nearest of 17 wild rice hot spot lakes with 100+ acres of predicted wild rice in 2018	Full ROI
Minimum Nearest Neighbor	Near Analysis; ArcGIS	The minimum euclidean distance of the nearest wild rice neighboring lake (out of the 10 nearest neighbors).	Full ROI
Average Nearest Neighbor	Near Analysis; ArcGIS	The average euclidean distance of the nearest wild rice neighboring lake (out of the 10 nearest neighbors).	Full ROI
Maximum Nearest Neighbor	Near Analysis; ArcGIS	The maximum euclidean distance of the nearest wild rice neighboring lake (out of the 10 nearest neighbors).	Full ROI
Connected Presence Lakes	Select by Location; ArcGIS	The number of lakes containing wild rice presence that are connected by immediate flowlines.	Full ROI
Reciprocal Area Hydraulic Load	NHDPlus	"Hydraulic metric of time required to displace one unit volume of water", (see Schmadel et al., 2018).	Lake
Arbolate Sum	NHDPlus	Cumulative distance of all upstream paths (U.S. Geological Survey, 2019).	Full ROI
Stream Order	NHDPlus	Strahler stream order	Full ROI
Path Length	NHDPlus	Distance to terminal flowline along the main path	Full ROI

Appendix 2B. Comparison between the top models from all three steps of the modeling framework using an iterative process in rfUtilities with human input (middle column) and the top models from the three-step modeling framework relying solely on algorithmic variable selection (right column).

Statistical Metric	Model generated via human input and algorithmic selection of predictors	Algorithmic selection of predictors
Random Forest	$R^2 = 0.359$ RMSE = 1.82	$R^2 = 0.332$ RMSE = 1.90
Linear Interaction Model	Adjusted $R^2 = 0.362$ AIC =1272.7	Adjusted $R^2 = 0.31$ AIC = 1303
Spatial Lag Model	AIC = 1266.4	AIC = 1292.5