Emily K. Bender
April 23rd, 2024

Wild Rice Detection using Synthetic Aperture Radar Imagery in Northwestern Minnesota

Introduction

The Ojibwe people settled in Minnesota after a migration from the East coast of the United States in the 1600s. They discovered wild rice in lakes and rivers and it became a dependable food source for them for a time. Ojibwe people harvested the wild rice in only as large quantities as they could then process right away. It was harvested in canoes- someone in a pair would row, while the other knocked the rice into the body of the canoe using sticks. Designated elders of the tribe would monitor the rice supply in the bodies of water that it existed and ensure that there were still some grains left unharvested for the next season, among other duties. In 1837, after the Treaty of Mendota, the US government seized Ojibwe land, making it very difficult for Ojibwe people to carry on with many traditional practices, such as wild rice harvesting (*MNopedia*, n.d.). Over time, non-natives began to harvest and process wild rice using new technologies and methods that endangered the rice stands. The hand-harvested rice became less desirable to the market due to cheaper options that came from cultivating wild rice. The plant is very responsive to changes in the water that it resides in, which could occur due to a number of reasons, some examples being climate change, and pollution.

Remote sensing turned out to be a useful tool in monitoring wild rice populations in nearby bodies of water for people such as the Ojibwe, who have a very personal, long-standing connection to wild rice as a food source. One piece of research that served as inspiration for this study, was one that combined the wealth of indigenous knowledge with multispectral imaging (Landsat-7 ETM+) of the water bodies of the Leech Lake Reservation in Minnesota to separate out what is wild rice vs other emergent aquatic vegetation such a bulrush and cattail (Price, 2012). Wild rice (scientific name: Zizania Palustris) exists in two varieties in Northern Minnesota, but I will not be attempting to tell them apart for the purposes of this paper. Wild rice can reach up to 6-9 feet above the surface of 1.5 to 3 feet of water. Wild rice stands can stretch for hundreds of hectares and become quite densely populated with wild rice alone. In his study, Michael Wassegijig Price tested four supervised classifiers for their ability to detect wild rice from open water and other aquatic vegetation: maximum likelihood, minimum distance to means, Mahalanobis distance and parallelepiped. Prior to testing the supervised classifiers, he tested several versions of an unsupervised ISODATA (Iterative Self-Organizing Data Analysis Technique) with anywhere from 2-6 maximum categories. Price found that the 5-category image performed best overall, and then assigned Boolean values (ones and zeros) to separate the upland category from all other categories in the 5-category ISODATA image to aid in the supervised classifiers job to separate land vegetation from water and aquatic vegetation. This masking made for more accurate supervised classification. The maximum likelihood and Mahalanobis distance algorithms both performed the best with a 79.03% accuracy of identifying the field data correctly.

RADAR stands for radio detection and ranging. There are both imaging and non-imaging radar sensors. Radar sensors are considered active sensors, emitting their own microwave pulses, and receiving them once they have made contact with an object on Earth's surface, returned as what is known as "backscatter". Microwaves can cut through cloud cover and canopy cover. This makes radar sensors incredibly useful in less than ideal weather conditions. There are several parameters that must be taken into consideration when using Synthetic Aperture Radar (SAR) data:

Wavelength: Longer wavelengths amount to more penetration through the target on Earth's surface. Different wavelengths correspond to different bands, making band selection highly variable depending on the goals of the remote sensing project.

Polarization: The radar antenna can transmit and receive vertically (V) or horizontally (H) polarized energy. Meaning that the signal can be HH (horizontally transmitted and received), VV, VH, or HV. Depending on the polarization of the signal, the sensor can pick up on certain horizontal and/or vertical properties of the target on Earth.

Backscatter: Surface roughness is in large part what makes for variations in color of a radar image. For example, rough surfaces will appear lighter in color, because much of the emitted energy from the sensor returns to it, rather than being scattered in all different directions. However, wavelength also has a part to play in the appearance of surface roughness. Height variations of the target must be smaller than the wavelength to appear smooth in the image (measured in centimeters).

Density of scattering targets on Earth will make for a stronger signal return if these objects are close together.

Dielectric constant has to do with the moisture content of a target. There is less penetration that can occur when there is more moisture within a target. Put differently, image brightness increases with greater moisture content in a target (NASA Video 1/4, 2018).

Efforts to predict northern wild rice vs. other emergent aquatic vegetation using SAR imagery have been fruitful, especially with Vertical-Vertical (VV) polarized imagery (O'Shea et al., 2020). O'Shea et al., generated 3 models that varied from model to model based on their definition of wild rice presence in percentages- the first model wild rice had to occur anywhere from 0-100% of the coverage in a given polygon, the second model made it so that wild rice made up anywhere from 30-100% of the vegetation stands in the designated polygon, making it the dominant or secondary taxa, the last model, used aspects of the second model, while also incorporating a VV threshold of >= -0.5.

Methods

I'd like to first acknowledge that my study area blankets the White Earth Reservation. This study is purely my own exploration of using SAR imagery to discern between aquatic emergent vegetation, specifically wild rice. I used only remote sensing data that was available and free to the public via NASA's EarthData, and the USGS EarthExplorer platform. For this study, I utilized Sentinel-1, C-band data. The frequency of the C-band is between 4-8 GHz, while the wavelength of the microwave is between 3.75 to 7.5

centimeters. The wavelength of the band does not allow for much penetration into the medium, but this is acceptable as that is not necessary for my particular research goals. Backscatter is the portion of the transmitted signal that is redirected back to the radar sensor. The SAR imagery I have incorporated into my model are both VV and VH polarized.

SAR is useful in the detection of wild rice especially once it has reached the emergence phase of the growing season. The radar signal will bounce either from the stalk of the rice to the surface of the water, or the water to the rice stalks, creating a double-bounce scatter pattern when returned to the sensor (Gallant et al., 2014). What does pose a challenge to a study like this is the variety of emergent aquatic vegetation (EAV) that exists in Minnesota bodies of water. For example, cattails and bulrushes both get categorized as medium to tall EAV depending on the time of the growing season, which is similar to wild rice. However, wild rice is unique in that its seeds get spread by the wind, rather than other EAV which tend to stay put for several consecutive growing seasons.

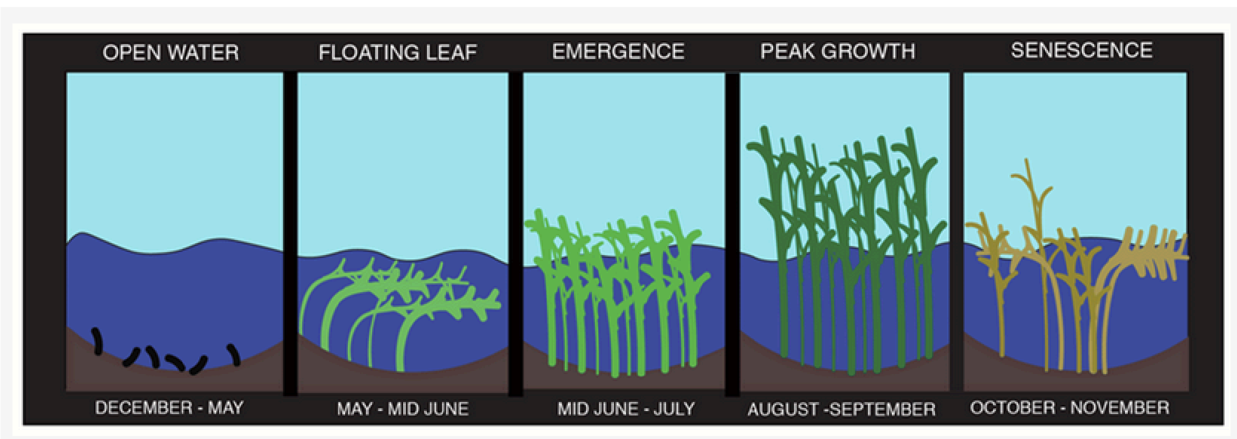

Figure 1. Annual life cycle of wild rice (O'Shea et al., 2020).

The study extent for my project falls within Synthetic Aperture Radar path 63, frame 151 and Landsat path 029, row 027. I collected SAR data from 06-19-2015, and Landsat data from 07-02-2015. I've chosen mid-June and July because this is the time of year when wild rice will start to peak out of the water, AKA the emergence phase. I converted all rasters to the WGS 1984 UTM, Zone 15N coordinate reference system. I clipped the rasters so as to get rid of all of the parts of the imagery that did not overlap. Using Landsat-8 imagery, I derived Normalized Difference Vegetation Index (NDVI) and Normalized Difference Moisture Index (NDMI) layers to add to my random forest algorithm as 'explanatory rasters'. I incorporated NDVI, as it measures plant health and greenness- NDVI values closer to 1 are indicative of healthy vegetation (dela Torre et al., 2021). NDMI, on the other hand, is used to monitor the water content in vegetation- NDMI values approaching one indicate no water stress, and total canopy cover (*NDMI*, 2022). These indices make use of several bands within multispectral data, applying various algebraic functions to draw out the features of vegetation on the ground (Baig et al., 2014).

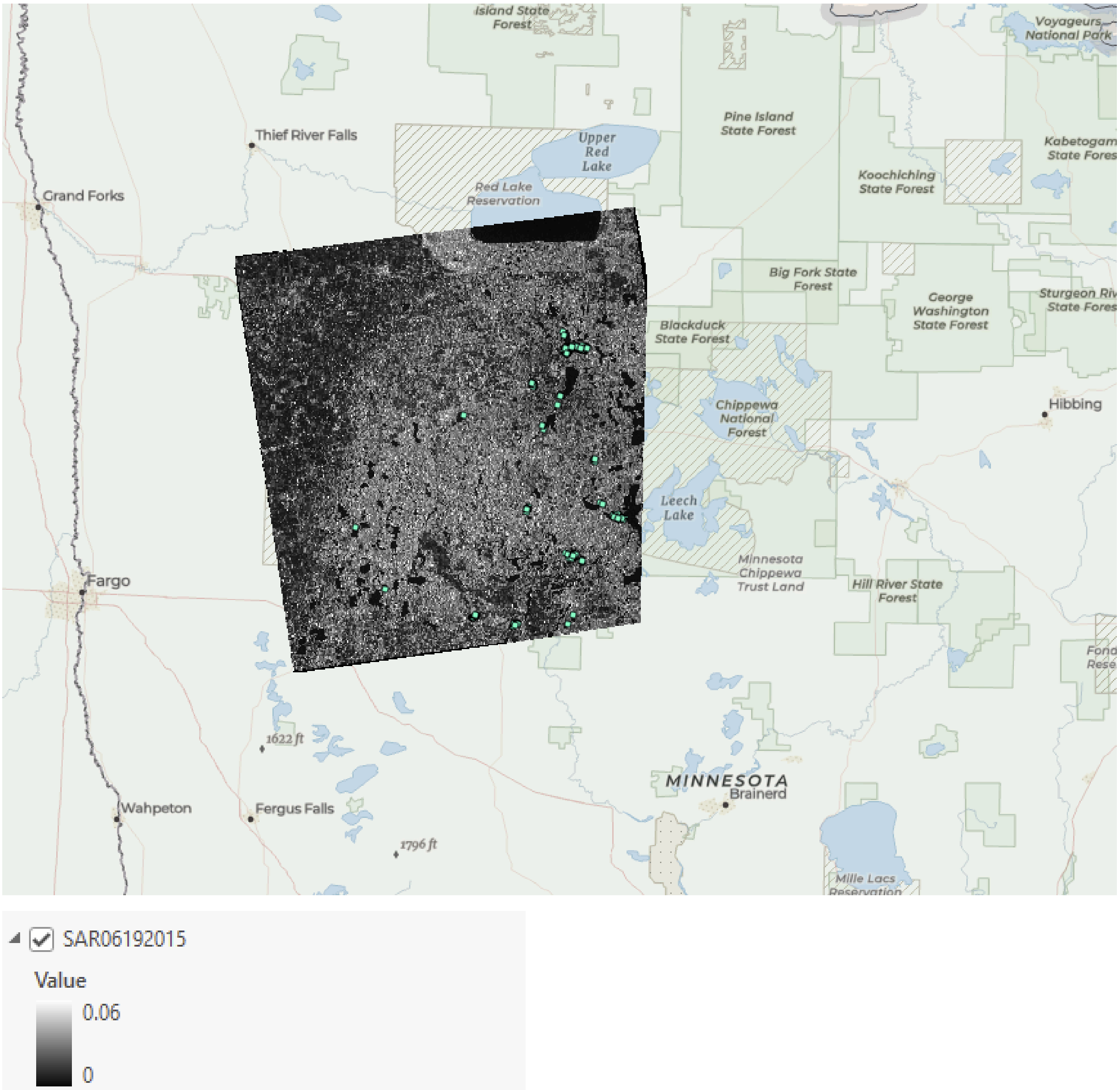

Figure 2. Preprocessed Synthetic Aperture Radar imagery clipped and brought into ArcGIS Pro to feed to the random forest algorithm, with DNR training data incorporated as points.

Once I've run these rasters through the random forest algorithm, they attempt to predict the presence of wild rice in several west central Minnesota bodies of water. I also have SAR data from 08-30-2015 that I ran through the random forest algorithm to see if there is a more robust prediction due to the plant being at its tallest height in the life cycle.

Through a contact with the Minnesota DNR, I was able to obtain a polygon shapefile with field data regarding the presence of various emergent aquatic vegetation (including wild rice). I created a binary, presence/absence field in ArcGIS Pro, so that wild rice existing in any capacity would be set equal to 1 and any other vegetation would be set equal to 0. From this, I derived points and clipped the layer to be fully contained by my study extent. This data is meant to be used as a reference of where these aquatic vegetation stands occurred throughout the state in the years leading up to 2015.

I used the SNAP toolbox to process my SAR imagery. I performed geometric and radiometric calibration according to a reference video I found produced by NASA (NASA Video 2/4, 2018). This preprocessing is an important step to reduce speckle in the SAR imagery, and derive as much of the information that is important to the study as possible and correct for any geometric distortions in relation to the actual XY coordinates of the target. Before running the SAR data through the random forest algorithm, I also converted the VV and VH aspects of the imagery from Sigma nought to decibels to allow for the measurement of backscatter.

Results

I chose to take a simpler approach to projects that I have referenced since I am relatively new to Synthetic Aperture Radar data, and running random forest algorithms. I took my presence/absence point layer, SAR image, NDVI and NDMI layers and fed it into a random forest algorithm of about 100 trees, saving 30% of the DNR training data points for validation of the model. The model was run in ArcGIS Pro and compensated for sparse categories which is helpful when you have a smaller field dataset to pull from, as was the case with my subset of data. I ran a "train only'' random forest algorithm twice to see how they would perform, as I made slight adjustments. For the first run, I used SAR imagery from mid-June along with NDVI and NDMI layers. All variables seemed to be equally important, but I felt that I wanted to incorporate a way to restrict where the model would predict wild rice presence. I then ran another "train only" model with an NDVI threshold applied to try and separate out land and water, this threshold turned out to be not so important to the predictive power of the model, so I dropped it for the final runs. I found that pulling a water mask into the random forest algorithm would work better.

I then ran the random forest algorithm using 'Predict to Raster' with the SAR imagery from mid-June, I added a water mask to aid the model's wild rice predictive power. I found a Hydrography dataset for the state on the Minnesota Geospatial Commons that I was able to clip to my area of interest and surround with a 5 foot buffer to account for changes that may occur to the water levels and shorelines. I then applied this dataset as a mask for my random forest algorithm to operate within and make predictions accordingly. Additionally, I derived three new rasters from Tasseled Cap Transformation (TCT) calculations: Wetness, Greenness, and Brightness. Brightness acts as a measure of overall reflectance, greenness is a measure of healthy vegetation and the density of its presence, and wetness further distinguishes grenness and brightness (*VEGETATION AND SOILS INFORMATION CONTAINED IN TRANSFORMED THEMATIC MAPPER DATA*, n.d.). After this, I noticed improvement to the accuracy of my random forest algorithm in predicting wild rice presence. I ran both the mid-June and late August SAR imagery through two different algorithms with the NDMI, TCT Brightness, Greenness, and Wetness variables and the water mask applied.

Based on the confusion matrix from the random forest algorithm the model using SAR imagery from the emergence phase of the growing season predicted that wild rice was present (true positive) with 53.9% accuracy. It predicted that wild rice was absent (true negative) with 80.4% accuracy. The model using SAR imagery from the peak growth phase, predicted wild rice was present (true positive) with 53.8% accuracy, while predicting that wild rice was not present (true negative) with 73.2% accuracy. We can also see that variable importance was quite similar between the two models.

Discussion

Model accuracy looks at how many times the classifier made correct predictions across the whole dataset. If we were just looking at model accuracy, then I would be in a pretty good position. However, accuracy values are not something that you can use to tell the whole story, especially when you have class imbalance. In those cases, the MCC and F-1 metrics become more valuable. This is the case for my study, because out of the 364 points I used to train my model, only 41 were wild rice presence points, which doesn't inform the model all that well what IS wild rice.

Matthews correlation coefficient (MCC) only approaches 1 when the model correctly predicts most of the positive and negative instances in the data, and if most of its predictions overall are correct (Chicco et al., 2021). That said, the MCC score is taking into account all four outcomes of the model: true positive, true negative, false positive, false negative. On the other hand, the F-1 score weighs precision and recall. The MCC metric and F-1 score can provide further insight into a sometimes misleading accuracy score. While my MCC values are not all that impressive for my final two raster outputs, they did improve from less than 0 and 0.06, in my "train only" models. This improvement occurred after I applied the water mask so that the model could only look at water features and anything within a 5 foot buffer of the water feature, and after I applied the tasseled cap transformation rasters to my collection of predictor variables. Once I made those changes, the model made less mistakes overall when predicting the presence or absence of wild rice in my study area.

Validation Data: Classification Diagnostics (06/19/2015 model):

Category	F-1 Score	Matthews Correlation Coefficient (MCC)	Sensitivity (True Positive Rate)	Accuracy
All	0.61	0.26	NA	0.77
0	0.86	0.26	0.80	0.77
1	0.36	0.26	0.54	0.77

Validation Data: Classification Diagnostics (08/30/2015 model):

Category	F-1 Score	Matthews Correlation Coefficient (MCC)	Sensitivity (True Positive Rate)	Accuracy
All	0.56	0.19	NA	0.71
0	0.82	0.19	0.73	0.71

1	0.56	0.19	0.54	0.71

Still, when looking at the two raster outputs from the random forest models, one from emergence and the other from peak growth, they look very different. Given these differences in appearance, I'd like to venture some educated guesses as to what could be causing this. In May of 2015, the Minnesota DNR released a hydrologic conditions report illustrating that the precipitation totals for the month were higher than their historical averages throughout the state. In many areas of the state, this average was surpassed by one to three inches ("Hydrologic Conditions Report - May 2015," 2015). This could have had an impact on the availability of ideal growing conditions for wild rice and provides further context to the environmental conditions of the time.
On some of the larger lakes, in the outer edges of my study, you will notice that there seems to be a lot of wild rice present in the emergence phase of the growing period, and this could be due to a misclassification of wild rice for other EAV, such as bulrush and cattails that also exist in these bodies of water. Needless to say, I do not believe all of the model's presence predictions to be wild rice.
There also appears to be a significant drop off between the presence of wild rice in the first model versus the model from late August. This could be due to harvesting taking place during the peak growth phase. I created a difference image to try and illustrate these changes, but it honestly wasn't as telling as I'd hoped it would be, so I did not include it in this paper.

If I were to continue with this study, I'd want to expand my study extent so that I can include more wild rice presence data into the model. I'd also be interested in working with the people who are indigenous to the area of study and incorporating some of their knowledge of the land, as well as working to see how this kind of research can be most beneficial to them. I'd also want to incorporate more variables into the model, such as areas of the lake where wild rice is most likely to be, or perhaps create a layer with the depths of the water bodies, looking for ideal growing conditions for the crop.

I really feel like I learned a lot throughout this process by reading other research papers and scouring through various resources. I had never run a random forest algorithm before this project, nor had I ever used the SNAP toolbox before. And I can confidently say that I know more than I did before the semester began. I don't know that I would be proud to turn this project over to a future employer yet, but it is a really good start and I feel like I can expand on this project in the future. While reviewing my peers' papers and research ideas I felt ill-equipped just because we are all coming at this with different goals and skills within the very big realm of remote sensing. However, I definitely appreciated the feedback that I received from my classmates throughout the process.

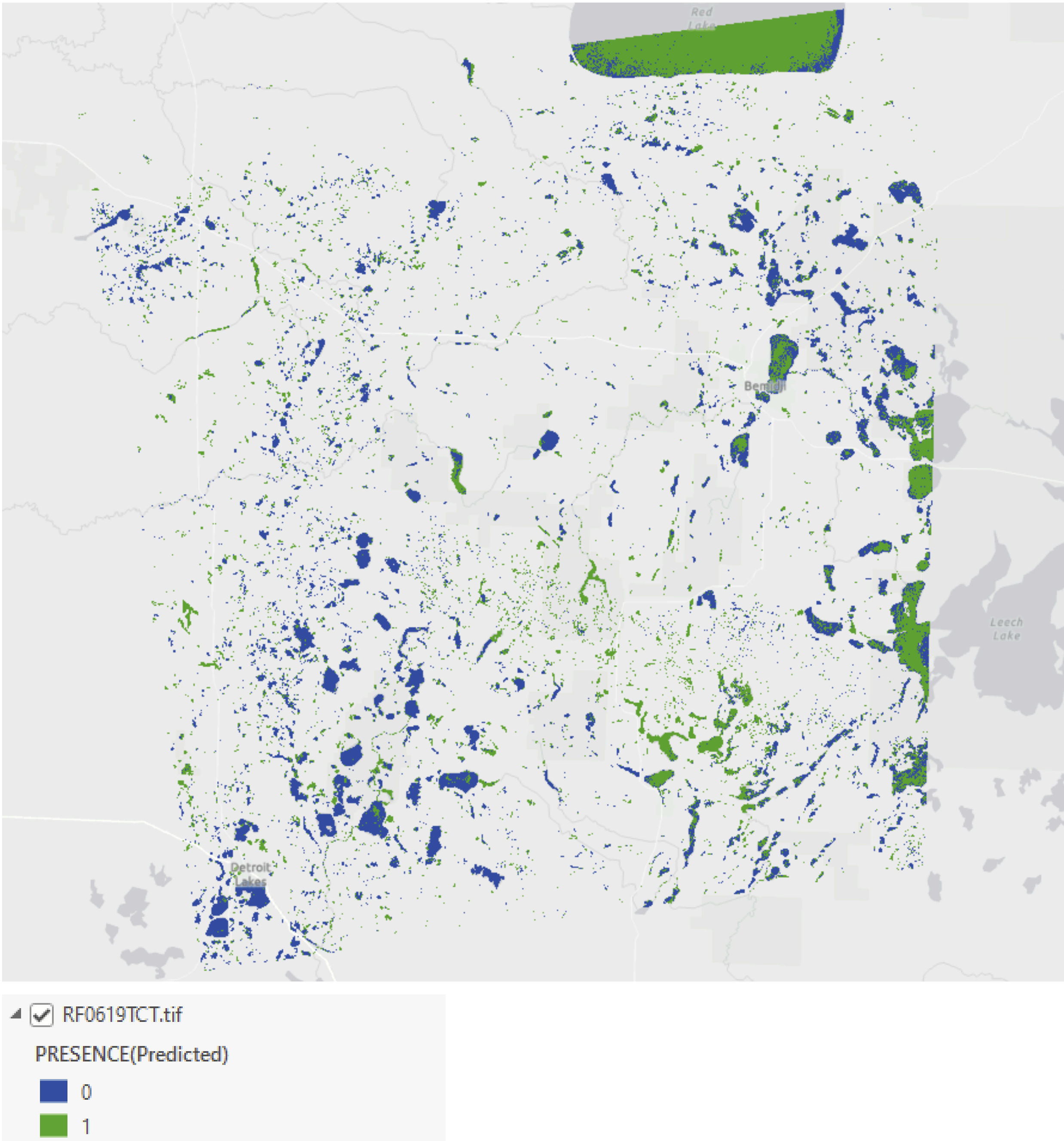

Figure 3. This is the output raster from the model with the SAR imagery from the emergence phase of the wild rice life cycle.

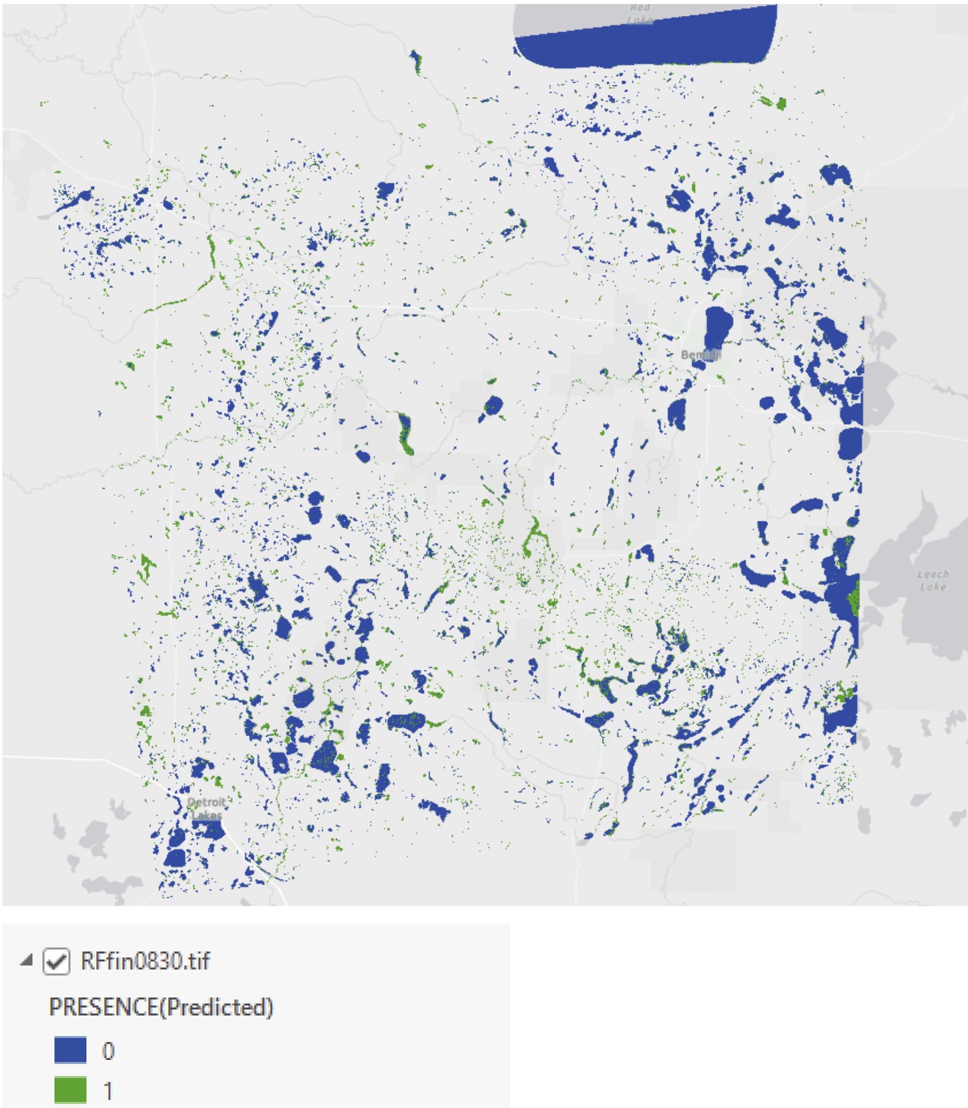

Figure 4. This is the output raster from the model with the SAR imagery from the peak growth phase of the wild rice life cycle.

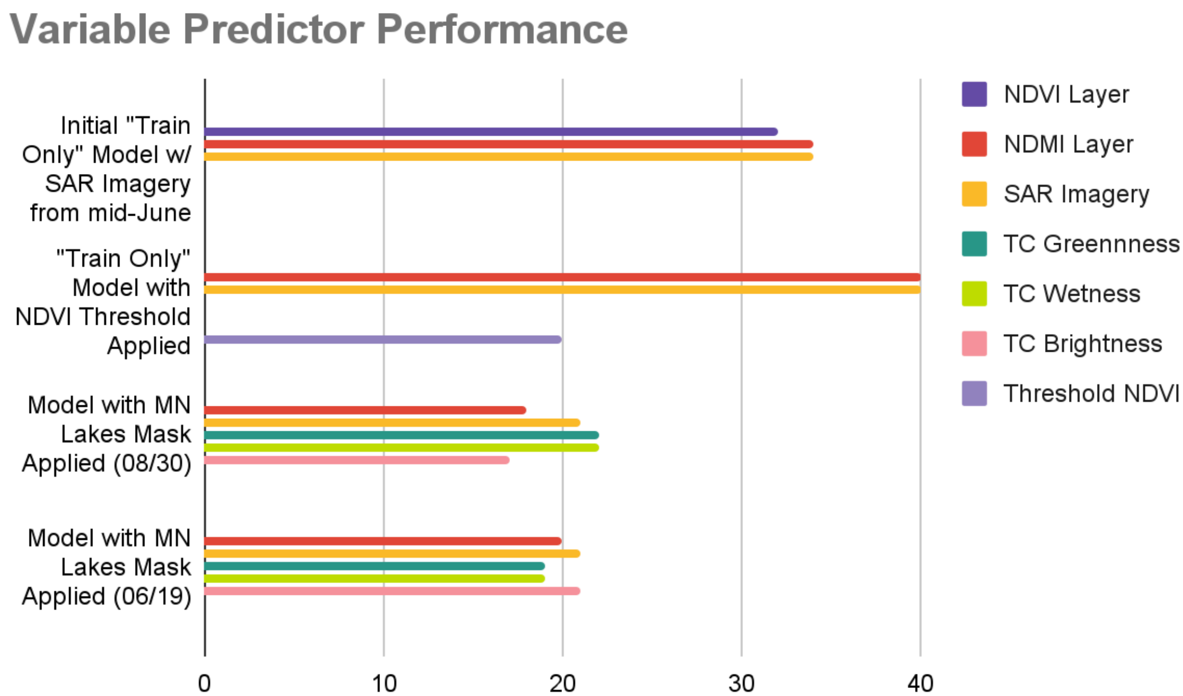

Figure 5. These are the results of train only and predict to raster random forest outputs with the variable importance of each model. I left the NDVI layer out of the final models because ArcGIS Pro would not allow me to run the algorithm with the layer being that there were too many categories.

References

- *Wild Rice and the Ojibwe | MNopedia*. (n.d.). Retrieved April 3, 2024, from https://www.mnopedia.org/thing/wild-rice-and-ojibwe
- Price, M. (2012). Spectral Identification of Wild Rice (Zizania palustris L.) Using Indigenous Knowledge and Landsat Multispectral Data. *Graduate Student Theses, Dissertations, & Professional Papers*. https://scholarworks.umt.edu/etd/910
- NASA Video (Director). (2018, August 23). *NASA ARSET: Basics of Synthetic Aperture Radar (SAR), Session 1/4*. https://www.youtube.com/watch?v=Xemo2ZpduHA
- Gallant, A. L., Kaya, 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), Article 3. https://doi.org/10.3390/w6030694
- dela Torre, D. M. G., Gao, J., & Macinnis-Ng, C. (2021). Remote sensing-based estimation of rice yields using various models: A critical review. *Geo-Spatial Information Science*, *24*(4), 580–603. https://doi.org/10.1080/10095020.2021.1936656
- *NDMI: Vegetation Index Equation And Values Interpretation*. (2022, February 7). https://eos.com/make-an-analysis/ndmi/
- Baig, M. H. A., Zhang, L., Shuai, T., & Tong, Q. (2014). Derivation of a tasselled cap transformation based on Landsat 8 at-satellite reflectance. *Remote Sensing Letters*, *5*(5), 423–431. https://doi.org/10.1080/2150704X.2014.915434
- O'Shea, K., LaRoe, J., Vorster, A., Young, N., Evangelista, P., Mayer, T., Carver, D., Simonson, E., Martin, V., Radomski, P., Knopik, J., Kern, A., & Khoury, C. K. (2020). Improved Remote Sensing Methods to Detect Northern Wild Rice (Zizania palustris L.). *Remote Sensing (Basel, Switzerland)*, *12*(18), 3023-. https://doi.org/10.3390/rs12183023

- NASA Video (Director). (2018, August 23). *NASA ARSET: SAR Processing and Data Analysis, Session 2/4*. https://www.youtube.com/watch?v=OwrLh7pjHRQ
- *VEGETATION AND SOILS INFORMATION CONTAINED IN TRANSFORMED THEMATIC MAPPER DATA*. (n.d.). Retrieved April 21, 2024, from http://www.ciesin.org/docs/005-419/005-419.html
- *DNR Hydrography Dataset—Minnesota Geospatial Commons*. (n.d.). Retrieved April 18, 2024, from https://gisdata.mn.gov/dataset/water-dnr-hydrography
- *ASF Data Search*. (n.d.). Retrieved April 18, 2024, from https://search.asf.alaska.edu/#/?polygon=POLYGON((-98.8384%2043.6075,-89.7648%2043.6075,-89.7648%2049.4333,-98.8384%2049.4333,-98.8384%2043.6075))&resultsLoaded=true&zoom=5.232&center=-100.851,40.362&start=2015-06-19T05:00:00Z&end=2015-08-19T04:59:59Z&granule=S1A_IW_GRDH_1SDV_20150818T001457_20150818T001515_007310_00A07E_1D47-METADATA_GRD_HD
- *EarthExplorer*. (n.d.). Retrieved April 18, 2024, from https://earthexplorer.usgs.gov/
- Chicco, D., Tötsch, N., & Jurman, G. (2021). The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Mining*, *14*(1), 13. https://doi.org/10.1186/s13040-021-00244-z
- Hydrologic Conditions Report—May 2015. (2015). *GR ANT*.