Minnesota Agriculture & Food Security

Implementing NASA Earth Observations to Validate Spectral Detection Models of Northern Wild Rice

 **Technical Report**

Final Draft – April 5th, 2019

Vanesa Martín (Project Lead)

Kristen Dennis

Leorah McGinnis

Eli Simonson

Dr. Paul Evangelista, Colorado State University, Natural Resource Ecology Laboratory (Science Advisor)

Dr. Catherine Jarnevich, United States Geological Survey, Fort Collins Science Center (Science Advisor)  
Tony Vorster, Colorado State University, Natural Resource Ecology Laboratory (Science Advisor)

Previous Contributors:

Katie Walker  
Charles Whittemore  
Jillian LaRoe  
Daniel Carver

# 1. Abstract

Crop wild relatives (CWRs) are genetically related to cultivated crops and function as repositories for genetic diversity. These plants have the potential to improve the yield, nutritional value, and resilience of crops, thereby buffering against widespread crop failure and supporting rural economic productivity. Our partners at the United States Department of Agriculture, Agricultural Research Service, National Plant Germplasm System (USDA ARS NPGS) are tasked with preserving CWRs, such as northern wild rice (*Zizania palustris* L.) in Minnesota. Additional partners at the Minnesota Department of Natural Resources (MN DNR) monitor northern wild rice and other aquatic vegetation by conducting annual field surveys. Currently, the NPGS relies primarily on habitat distribution modeling to predict suitable habitats for CWRs, while the MN DNR relies on its field surveys. This project focused on validating a digital ocular sampling (DOS) method for wild rice detection with the intention of improving habitat distribution modeling and facilitating wild crop monitoring. Previous NASA DEVELOP research incorporated Sentinel-1 C-band Synthetic Aperture Radar (C-SAR) and Landsat 8 Operational Land Imager (OLI) into a DOS approach, which was subsequently validated in this project with *in situ* data collected by the MN DNR. Results confirmed that models trained with MN DNR field data, as well as a combination of spectral and radar variables, outperformed other model iterations and are highly accurate in their classification of rice. We also found that DOS is not suitable for training models to distinguish between different aquatic vegetation types. These insights will provide end users with an improved framework for integrating remote sensing into their wild rice monitoring.

**Keywords**

*Zizania palustris* L., crop wild relative, digital ocular sampling, Landsat 8 Operational Land Imager (OLI), Sentinel-1 C-Band Synthetic Aperture Radar (C-SAR), random forests

# 2. Introduction

***2.1 Crop Wild Relatives***

Crop wild relatives (CWRs) are plants that are genetically related to cultivated crops and function as repositories for genetic diversity (United States Department of Agriculture Forest Service, 2014). They also comprise a significant portion of threatened native plant species across the United States (Meilleur & Hodgkin, 2004). Approximately one in every five plant species across the world is at risk of extinction; threats include habitat degradation, invasive species, pollution, and agricultural modernization (Brummitt & Bachman, 2010). The conservation of CWRs both in their natural habitat (*in situ)* and in seed banks (*ex situ*)is critical to safeguard genetic diversity and prevent species loss, especially for staple crops (Khoury et al., 2013). Domestication of crops contributes to both phenotypic and genetic alteration, which may cause species to lose important traits necessary for survival in their natural habitats (Suszkiw, 2014). Conserving CWRs *in situ* allows for the genetic preservation of these populations and subsequent species within their natural habitat, and these unique genetic traits can be reintroduced to bolster strength, resilience, yield, and nutritional value of domestic crops (Khoury et al., 2013).

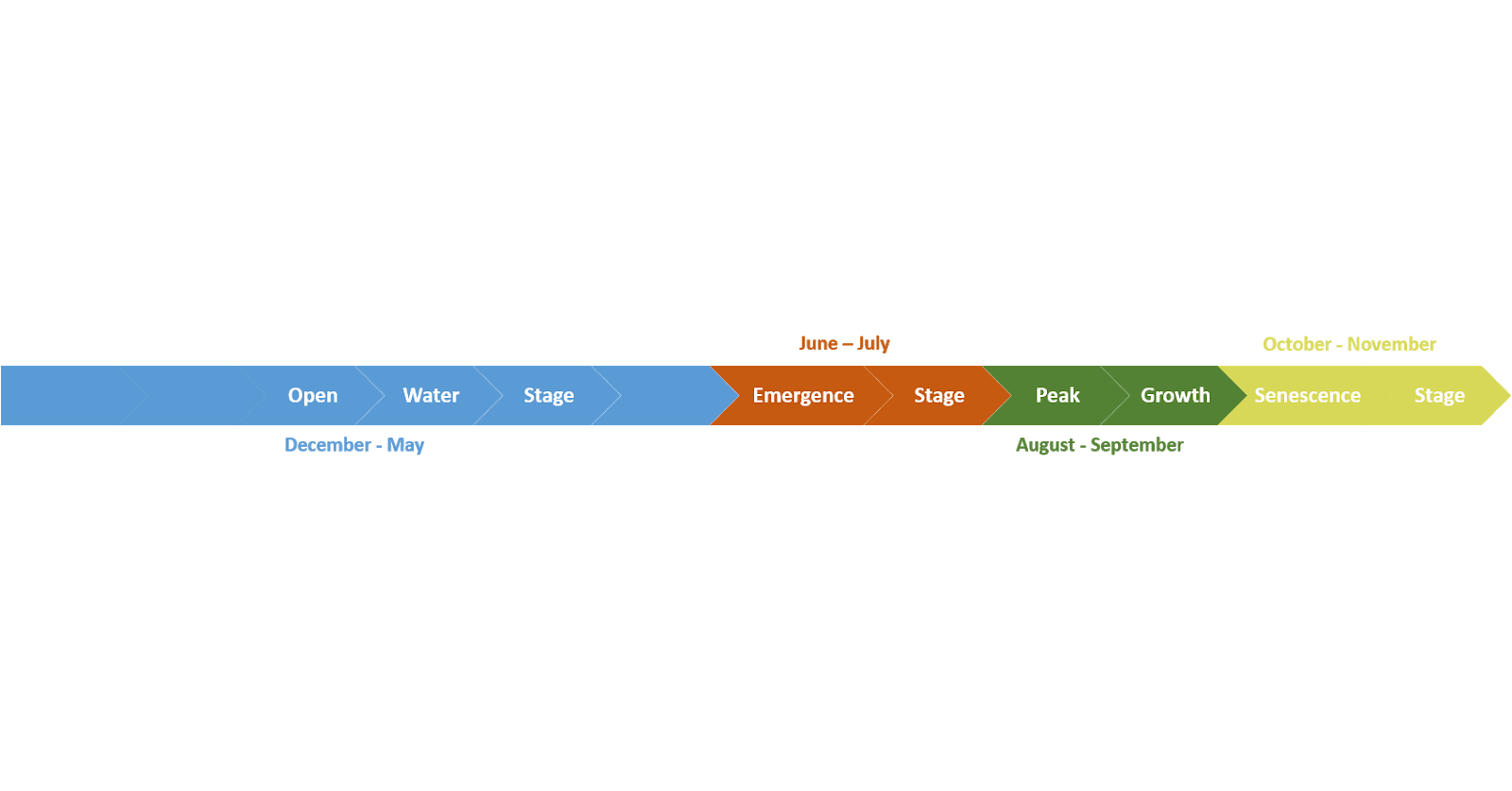


*Figure 1.* Northern wild rice (*Zizania palustris* L.) in Lake Itasca (photo by Jillian LaRoe).

***2.2 Focal Species***

Northern wild rice (*Zizania palustris* L.) (*Figure 1*) has been harvested in the Great Lakes region for about two thousand years (Minnesota Department of Natural Resources, 2008). The Ojibwe tribe of the northern Great Lakes region has a strong cultural connection to wild rice. Their migration from the east coast was guided by a mission to find “food on water,” or *manoomin,* which led them to Minnesota’s lakes and rivers that support emergent aquatic wild rice (Minnesota Department of Natural Resources, 2008). Besides being an important part of the cultural identity of indigenous peoples, it also represents a significant source of economic development for the region as Minnesota is one of the nation’s leading producers of wild rice (Minnesota Department of Natural Resources, 2008).

Wild rice is an annual aquatic plant that grows in monotypic stands in freshwater bodies, wetlands, and along stream channels (Price, 2012). Stands of wild rice grow in non-stagnant water depths between 0 and 1.5 meters and can grow up to 100 hectares in area (Price, 2012). Stand density and size vary year-to-year and are highly susceptible to changes in water levels and water chemistry (Minnesota Department of Natural Resources, 2008). Northern wild rice has unique seasonal phenology that can be summarized into four phases for remote sensing purposes. In the open water stage (December to late May), northern wild rice is submerged beneath the water. During the emergent stage (June to July), the rice grows horizontally along the water’s surface until it is rigid enough to emerge and stand vertically. At the peak growth stage, the plant matures to produce seeds (August to early September). During this stage, the plant can be up to 1 meter in height and is traditionally harvested (Minnesota Department of Natural Resources, 2008). During the senescence stage (October to November), rice stands lose their green color and eventually fall into the water. Understanding the phenological development of wild rice allowed us to identify specific time periods when remote sensing is more likely to detect wild rice (*Figure 2*) and utilize imagery representative of these time periods to differentiate this species from other aquatic vegetation.



*Figure 2.* Yearly timeline of northern wild rice growth.

Previous studies have successfully mapped rice species using Landsat 5 Thematic Mapper (TM), Landsat 8 Operational Land Imager (OLI), and Sentinel-1 C-Band Synthetic Aperture Radar (SAR) remote sensing data (Brandt, 2008). SAR imagery has proved useful in identifying the unique spectral signature of wild rice due to the sensor’s ability to transmit through cloud cover and detect changes in surface roughness. Nelson et al. (2014) found SAR more effective at monitoring rice over time in Asia than relying on spectral imagery alone. Datasets such as bathymetry and hydrology have also been used as a means to define suitable growing habitat for wild rice (Anderson, Kapfer, Ueland, Lawrence, & Cordts, 2011; Brandt, 2008).

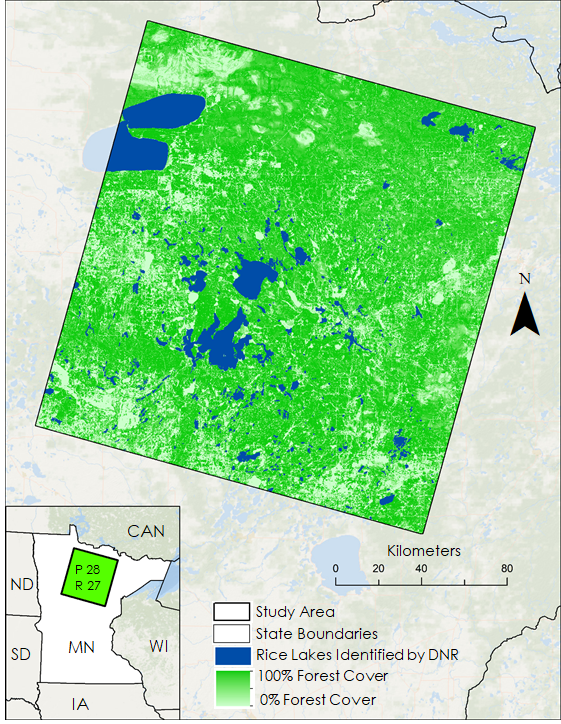
***2.3 Project Partners & Objectives***

The Colorado – Fort Collins Node continued its partnership with the United States Department of Agriculture Agricultural Research Service National Plant Germplasm System (USDA ARS NPGS), an organization with the primary goal to preserve genetic diversity for the security of our food resources in the future.For this reason, it is critical to understand the extent and range of specific CWRs throughout the US. A related DEVELOP project previously employed NASA Earth observations to create models detecting emergent aquatic vegetation in order to aid the USDA in determining the extent of northern wild rice. The research described here used *in situ* data collected by partners at the Minnesota Department of Natural Resources (MN DNR) to validate digital ocular sampling (DOS) models of emergent aquatic vegetation and determine the feasibility of using satellite imagery for distinguishing wild rice from other aquatic vegetation. The University of Minnesota provided social and biological insights into our species that were vital to this project.

The goal of this study was to provide partners with an optimized methodology for detecting emergent aquatic vegetation in Minnesota, including northern wild rice (*Zizania palustris* L.), that is applicable beyond a single location and/or time period. We aimed to accomplish this by using *in situ* data collected by partners at the MN DNR to validate DOS models of emergent aquatic vegetation and refine the use of satellite imagery for distinguishing wild rice from other aquatic vegetation. This improved methodology enables our partners at the USDA ARS and MN DNR to more effectively incorporate NASA Earth observations into their mission and support field monitoring.

***2.4 Study Area and Period***

Our main study area and period covered a single Landsat scene (Worldwide Reference System 2, Path 28, Row 27) in North Central Minnesota (*Figure 3*) and spanned June to September 2015. This extent spans roughly 34,000 square kilometers and ranges from 330 to 560 meters in elevation. The landscape is home to a vast network of rivers and lakes, and the unique wetland habitats across the study area create ideal growing conditions for northern wild ricepopulations (Minnesota Department of Natural Resources, 2008). Specifically, we looked at lakes surveyed by the MN DNR within this scene. Sentinel-2 MultiSpectral Instrument (MSI) data were not available for a majority of this extent and time period. Therefore, we conducted a separate case study in another area of north-central Minnesota for June to September 2017 where Sentinel-2 MSI, Sentinel-1 C-SAR, and MN DNR *in situ* data were all available (Appendix A).



*Figure 3.* This is the study area extent, which is situated in north-central Minnesota (Landsat Path 28, Row 27).

# 3. Methodology

***3.1 Data Acquisition and Processing***

***3.1.1 Species Presence Data***

Presence data for northern wild rice and co-occurring aquatic vegetation were provided by the previous NASA DEVELOP spring 2018 Minnesota & Texas Agriculture team. These points were collected remotely through an approach known as DOS, where northern wild rice presence and absence points were opportunistically placed over one-meter resolution 2015 aerial imagery from the National Agricultural Imagery Program (NAIP) in Google Earth Engine (GEE) (Gorelick et al., 2017). Presence points were defined as locations where a Landsat pixel (30 m x 30 m) was estimated to contain at least >50 percent wild rice cover. Absence points were selected at locations that contained no wild rice and were sampled equally from water and other types of terrestrial vegetation that surrounded lakes. A total of 1,125 presence/absence points were generated for the state of Minnesota. We subset these points to our study area extent.

*In situ* polygon data for northern wild rice and co-occurring aquatic vegetation were provided by our collaborators at the MN DNR in the form of polygon delineations. These polygons identified the dominant taxa observed, as well as all other taxa within the same area. All species presence data were reprojected into the WGS 1984 UTM 15N coordinate reference system, rasterized, aligned to our Landsat scene, and clipped to the study area extent. Rasters were then converted to points and randomly subset to obtain separate model training and validation points. Mixed vegetation polygons containing wild rice were classified as presence points, regardless of the dominant taxa observed. 500 presence points and 500 absence points were generated for each set of points (training and validation). Initial model results based solely on MN DNR training data classified water as wild rice, so we placed an additional 200 absence points over open water in both our main study area and the Sentinel-2 case study area in order to better represent water in these models. The same validation dataset was used across the three models run over our study area and the model run in our case study area. Based on the MN DNR survey availability in 2015, the presence locations used to guide and validate this study were predominantly concentrated in the southern half of our main study area.

***3.1.2 Satellite-Derived Spectral Data***

All satellite imagery and datasets were acquired and processed using Google Earth Engine. Tier 1 Top of Atmosphere (TOA) reflectance, surface reflectance, vegetation and moisture indices (Normalized Difference Vegetation Index [NDVI], Normalized Difference Moisture Index [NDMI]), and tasseled cap components (wetness, brightness, and greenness) were derived from Landsat 8 OLI and Sentinel 2 MSI (Appendix Tables B1 and B2). Sentinel-1 extra wide swath C-Band SAR imagery was processed to obtain radar backscatter (Appendix Tables B1 and B2). When processing satellite imagery, we considered the distinctive traits of our focal species’ growth cycle, such as its emergence from water surfaces during the summer. All image collections were then filtered by the summer (June to July) and fall (August to September) seasons of 2015 for our Landsat scene and 2017 for our Sentinel case study area. A cloud filter was then applied to obtain Landsat 8 OLI imagery with 35 percent or less cloud cover and Sentinel-2 MSI imagery with 5 percent or less cloud cover. An image composite for each season was produced by taking the median value at each pixel for all multispectral and radar bands.

Median composite TOA bands were used to produce spectral indices and tasseled cap components. Median composite radar bands were used to calculate radar ratios, range, and variance (Appendix Table B2). Previous NASA DEVELOP research showed that radar variance was a significant predictor variable, and setting a threshold value for variance yielded a conservative but accurate delineation of northern wild rice (Walker, Carver, LaRoe, & Whittemore, 2018). Consequently, we applied numerous thresholds, which we refer to as radar masks, to radar variance. After we found that radar range was the most significant predictor variable for our refined models, we also created radar masks with range and validated both masks in this research.

***3.2 Data Analysis***

***3.2.1 Validation Methods***

To validate the detection model from the previous DEVELOP Spring 2018 term that was trained using DOS presence data, we overlaid validation presence and absence points from MN DNR *in situ* data. We then extracted the model predictions to the validation points, with 0 being absence and 1 being presence. Finally, we compared the observed absence and presence to the predicted absence and presence to determine classification accuracies. Misclassified vegetation was further broken down by vegetation type to determine which aquatic vegetation classes were most frequently confused for our focal species. This validation method was applied to the preliminary DOS model and radar mask, as well as refined masks and models that were trained with MN DNR data.

***3.2.2 DOS Analysis***

To determine the feasibility of improving current DOS methods, we assessed whether 1 m NAIP imagery could be used to distinguish wild rice from other types of co-occurring aquatic vegetation, including cattails, bulrushes, and waterlilies. MN DNR polygon delineations for each vegetation type were imported into GEE and overlaid onto NAIP imagery. A visual assessment was performed for each vegetation class to determine the feasibility of accurately distinguishing them over aerial imagery.

***3.2.3 Modeling Methods Step-by-Step***

We employed the random forest (RF) classification model, developed by Breiman (2001), in this study. This model combined species presence and absence data with satellite-derived environmental variables (spectral reflectance, radar backscatter, and et cetera) to produce a binary classification of species absence/presence across our study area for northern wild rice. While RF models can be limiting because they require true-presence and true-absence data, they are also generally known to produce more robust outputs when compared to other classification models (Stohlgren et al., 2010).

The values for each predictor variable across all species presence and absence locations were extracted in GEE and exported as a CSV file. The Variable Selection Using Random Forest (VSURF, version 1.0.3) library in R studio was used to determine which spectral predictor variables had the most explanatory power for distinguishing our focal species. From the CSV, the top predictors were ranked in order of explanatory power, and highly correlated variables (|r|>0.75) were removed. Models were run in GEE using the variables defined by the variable selection process. We processed four different models with varying sets of predictor variables (Table 1). In addition to confusion matrices, the following model evaluation metrics were calculated: area under the curve (AUC), sensitivity, and specificity (Appendix Table D1).

Table 1

*A description of the model iterations conducted in this study*

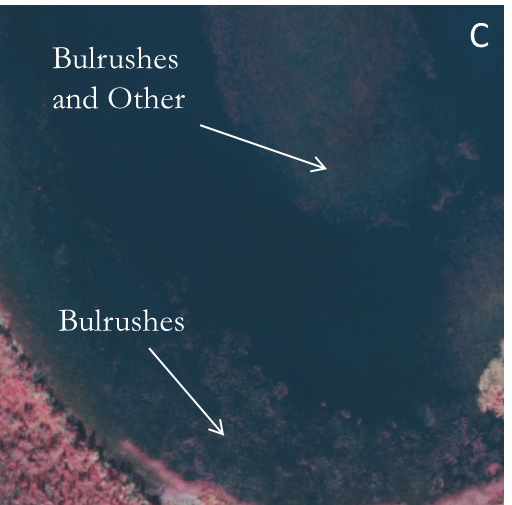
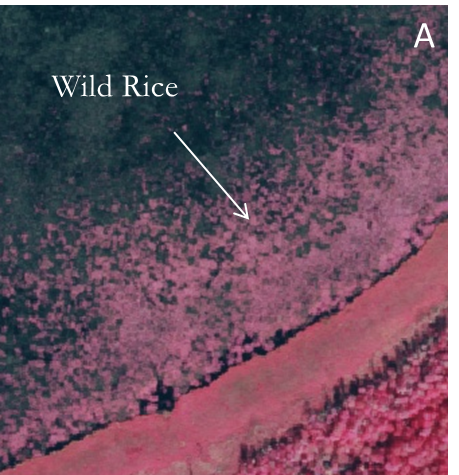
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model # | Training Data Used | Satellites/Sensors Used | Resolution (meters) | Top 4 Predictors |
| 1 | DOS | Landsat 8 OLI  Sentinel-1 C-SAR | 30 | VV range, Cross Ratio (VH/VV) T2, Cross Ratio (VV/VH) T2, Band 5 T2 |
| 2 | MN DNR | Landsat 8 OLI  Sentinel-1 C-SAR | 30 | VV range, Tasseled Cap Wetness T1, VV median T1, Band 2 T1 |
| 3 | MN DNR | Sentinel-1 C-SAR | 10 | VV range, VH range, VV median T1, Cross Ratio (VH/VV) T1 |
| 4 | MN DNR | Sentinel-1 C-SAR  Sentinel-2 MSI | 10 | VV median T1, VV range, Band 2 T2, Band 2 T1 |

# 4. Results & Discussion

***4.1 Validation of DOS Model***

Model 1 represents the DOS-trained model, and while it correctly classified over half of the MN DNR’s wild rice survey, it also misclassified 54 percent of other vegetation as wild rice (Table 2). The misclassification was likely due to a lack of absence points placed over co-occurring aquatic vegetation, meaning that the model was not trained to differentiate between rice and other aquatic vegetation. The model’s comparatively high AUC value of 0.94 (Appendix Table D1) can be attributed to the fact that the metric was calculated using the original DOS before we conducted an independent validation of the model using MN DNR field data.

***4.2 Digital Ocular Sampling Evaluation***



*Figure 4.* Examples of aquatic vegetation as seen through NAIP 2015 aerial imagery

A) wild rice, C) bulrushes and other

Attempting to improve the DOS sampling methodology from the spring 2018 iteration of this project led to some observations about the appearances of aquatic vegetation over high-resolution imagery. Wild rice typically has a spongy texture, but its density and coloring can vary widely. Similarly, waterlilies can appear spongy but can vary from pink to red, lending themselves to be frequently confused for rice. Bulrushes commonly appear as faint and translucent patches near lake edges, making them difficult to detect (*Figure 4*). Overlaying MN DNR vegetation polygons over aerial imagery highlighted the difficulties of distinguishing wild rice from other aquatic vegetation, particularly waterlilies and bulrushes. While there may be particular areas where these vegetation classes are distinguishable, these trends do not hold across larger geographic extents. For example, Appendix C shows that waterlilies and wild rice appear nearly identical in NAIP imagery despite the fact that they have been identified in the field as different species by the MN DNR. Generally, aquatic vegetation classes could not be identified or distinguished with confidence, making DOS a suboptimal method for training these detection models.

***4.3 Model Validation***

Table 2

*Validation results from DOS- and DNR-trained models with different predictor variable combinations*

|  |  |  |  |
| --- | --- | --- | --- |
| Model # | % Wild Rice Classified as Wild Rice | % Other Vegetation Classified as Wild Rice | Breakdown of Misclassified Vegetation |
| 1 | 68 % | 54 % | Bulrushes – 67 %  Waterlilies – 21 %  Cattails – 8 % |
| 2 | 81 % | 22 % | Bulrushes – 58 %  Waterlilies – 25 %  Cattails – 7 % |
| 3 | 67 % | 25 % | Bulrushes – 59 %  Waterlilies – 25 %  Cattails – 9% |
| 4 | 77 % | 20 % | Bulrushes – 80 %  Waterlilies – 19 % |

Models 1 and 2 were both run with 30 m Landsat and SAR variables. However, Model 2 was trained with MN DNR data rather than DOS; notably, it was the top performing model with misclassification reduced from 54 percent in Model 1 to 22 percent and correct classification boosted from 68 percent to 81percent. Across the board, models trained with the MN DNR’s *in situ* data performed better than model trained with DOS data. Furthermore, given that radar variables tended to be well-represented as significant predictor variables across model iterations, we opted to run another *in situ* data-trained model exclusively with 10 m C-SAR variables (Model 3); it also outperformed the DOS-trained model (Model 1) but did not accurately detect as much rice as Model 2. Through this finding, we determined that an *in situ* data-trained model run with a combination of spectral and radar variables is optimal for detecting and distinguishing wild rice, as opposed to one trained with DOS data or run with radar-based variables alone.

Further research is necessary to determine exactly why a combination of spectral and radar variables produced the most accurate detection output. The top spectral predictors for Models 1 and 2 included Landsat 8’s Band 2, the blue band, and Band 5, the near-infrared band. The blue band is generally employed for bathymetric mapping as well as distinguishing soil from vegetation. The near-infrared band can measure and detect biomass content as well as shorelines. Lastly, the wetness component of the Tasseled Cap transformation was also a top predictor for Model 2 as this variable is able to detect moisture content (Barsi, Lee, Kvaran, Markham, & Pedelty, 2014). All of these applications are relevant to detecting and distinguishing aquatic vegetation within lake shorelines. It is highly possible that taking these environmental factors into consideration, rather than focusing on radar signatures exclusively (*Figure 5; Appendix Figures E1 to E4*), led to a higher rate of accurate rice detection and classification.

Lastly, Model 4 represents our Sentinel-2 case study model carried out over a different area in north-central Minnesota over June through September of 2017 rather than 2015. The model performed strongly, misclassifying only 20 percent of other vegetation as wild rice. This robust performance may be due to the model’s combination of spectral and radar variables at 10 m, a higher resolution than Model 2 predictor variables. However, because the model was trained with different data over a different geographic extent, this explanation is not conclusive and the model is not directly comparable to the first three that were run.

Table 2 also indicates which aquatic vegetation types were most frequently misclassified as wild rice, and the consistent pattern throughout all model iterations is that bulrushes dominate this category, followed by waterlilies and cattails, respectively. On average, bulrushes represented 60 percent of the other aquatic vegetation that was misclassified, while waterlilies represented approximately 25 percent. Regarding Model 4, the misclassification breakdown was even more drastic, with bulrushes making up 80 percent and waterlilies comprising 19 percent of the vegetation types most commonly misclassified for wild rice. It is interesting to note that bulrushes and wild rice have a comparable radar signature relative to other aquatic vegetation which could explain why there is a larger level of confusion occurring (*Figure 5; Appendix Figure E2*). Waterlilies are found to grow alongside rice very frequently given their similar environmental preferences (Myrbo et. al, 2017), and co-occurrence was common in the MN DNR data. Our definition of wild rice presence encompassed mixed vegetation classes where rice was not a dominant species, and it is likely that this accounts for why waterlilies were misclassified for wild rice less frequently than bulrushes were.

# 

*Figure 5.* The radar signature of wild rice from April through November 2015.

Ultimately, it appears that the surveying methods employed by the MN DNR are conducive to training detection models. The caveat to this is that MN DNR data do not include absence points over open water, but those were easily collected and added to the dataset so that our detection models could differentiate between water and vegetation. Once those points were added, the MN DNR’s existing monitoring program proved to be well-suited for detection.

***4.4 Radar Mask Validation***

Table 3

*Results from testing the ability of threshold radar variance and range masks to detect northern wild rice based on 1,000 MN DNR validation data points*

|  |  |  |
| --- | --- | --- |
| Radar Mask | % Wild Rice Classified as Wild Rice | % Other Vegetation Classified as Wild Rice |
| Range ≥ 4 backscatter units | 50 % | 12 % |
| Variance ≥ 10 backscatter units | 48 % | 21 % |
| Range ≥ 7 backscatter units | 29 % | 1.5 % |
| Variance ≥ 30 backscatter units | 15 % | 1 % |

As mentioned earlier, radar variance and radar range were significant predictor variables and therefore thresholders to create a radar mask for wild rice. An analysis of the different variance and range thresholds showed that setting thresholds for radar range rather than variance results in higher classification accuracy and lower misclassification of other aquatic vegetation (Table 3). Validation of the spring 2018 iteration of this project’s wild rice mask (radar variance ≥ 30) showed that 15 percent of wild rice was correctly identified and only 1 percent of other aquatic vegetation was misclassified as wild rice. In direct comparison to the conservative variance threshold, the range threshold yielded nearly double the amount of correctly classified rice (29 percent) while keeping the misclassification of other aquatic vegetation down to 1.5 percent (Table 3). For both radar variance and range, as the threshold value is lowered, more wild rice is identified but other vegetation is misclassified at a higher rate. This thresholding approach can be used alone or in combination with modeling methods to produce informed understandings of wild rice distributions.

***4.5 Future Work***

There is potential interest from our collaborators at the Minnesota Department of Natural Resources to develop a Google Earth Engine tool that can be used to detect and monitor wild rice. Additionally, finding a geographic extent and time period where NAIP, Landsat 8, Sentinel-2, Sentinel-1, and MN DNR data are all available would improve ease of comparison. In this 10-week project, we were unable to overcome issues presented by this compound data availability. Specifically, we did not conduct a direct comparison in the same year between Sentinel-2 and Landsat 8 based models. Future work could also be done to determine how well these methods perform in other years, as well as to reduce misclassification of other emergent vegetation with different combinations of radar masks and detection models.

# 5. Conclusions

Our research yielded insights into various aspects of the detection of emergent aquatic vegetation. Our initial aim to compare models trained with DOS to models trained with *in situ* MN DNR data indicated some key conclusions. The first, that not all emergent aquatic vegetation is distinguishable using high-resolution NAIP imagery and that bulrushes and waterlilies are commonly misclassified as wild rice due to their spectral similarities. We also learned that training a model with digital ocular sampling was not optimal for distinguishing northern wild rice while MN DNR monitoring data did prove suitable for creating northern wild rice detection models. Additionally, results indicated when detecting northern wild rice radar range is a powerful variable. Although rice has a unique and reliable radar signature, we can detect rice with higher accuracy by combining spectral and radar variables rather than employing radar exclusively. Lastly, findings indicated that higher resolution variables do not necessarily result in higher detection accuracy. Understanding the limitations of DOS for training data and the benefits of combining spectral and radar variables to filter imagery for wild rice may be able to help partners continue to improve upon these models for more effective monitoring. Overall, these findings provide important insights into the potential of remote sensing to be useful in monitoring northern wild rice and emergent aquatic vegetation.

# 

# 6. Acknowledgments

The Minnesota Agriculture & Food Security team would like to thank our partners at the USDA, the University of Minnesota, and the MN DNR for engaging with NASA DEVELOP to make this research possible. We would also like to give a special thanks to Kristen Dennis (Acting Center Lead of the DEVELOP Colorado – Fort Collins and Idaho – Pocatello Nodes), Tony Vorster (Science Advisor), and Dr. Colin Khoury (USDA ARS NPGS) for working closely with our team throughout this research project.

This material includes modified Copernicus Sentinel data (2015 & 2017), processed by ESA.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Aeronautics and Space Administration.

This material is based upon work supported by NASA through contract NNL16AA05C.

# 7. Glossary

**ARS** – Agricultural Research Service

**AUC** – Area Under Curve:A statistic obtained via the Receiver Operating Characteristic (ROC) curve that is used for model evaluation and comparison, which describes the predictive power of a suite of environmental variables relative to presence and absence points, ranging from 0 to 1

**CWR** – Crop Wild Relative

**Earth observations** – Satellites and sensors that collect information about the Earth’s physical, chemical, and biological systems over space and time

**GEE** – Google Earth Engine, a cloud environment for large scale satellite imagery processing and analysis

***In situ* conservation** – “On-site” conservation; the process of protecting an endangered plant or animal species in its natural habitat

***Ex situ* conservation** –“Off-site” conservation; the process of preserving components of biological diversity outside their natural habitats (i.e. in seed banks or other genetic repositories)

**Monotypic** **stands** – Vegetation stands having a single form or member, especially containing no more than one taxonomic category

**NAIP** – National Agricultural Imagery Program

**NDVI** – Normalized Difference Vegetation Index

**NDMI** – Normalized Difference Moisture Index

**NPGS** – National Plant Germplasm System

**RF** – Random Forest classification

**Sensitivity** – A statistic that measures the proportion of positives (presences) correctly identified

**Specificity** – A statistic that measures the proportion of negatives (non-presences) correctly identified

**VSURF** – Variable Selection Using Random Forest

# 8. References

Anderson, R., Kapfer, P., Ueland, J., Lawrence, J., & Cordts, S.(2011).Comparison of wild rice and waterfowl surveys [PDF file]. Retrieved from [https://www.lccmr.leg.mn/projects/2011/finals/2011\_04j\_2e\_rpt\_wild-rice-data -waterfowl-surveys.pdf](https://www.lccmr.leg.mn/projects/2011/finals/2011_04j_2e_rpt_wild-rice-data-waterfowl-surveys.pdf)

Barsi, J.A., Lee, K., Kvaran, G., Markham, B.L., & Pedelty, J.A. (2014). The spectral response of the Landsat-8 Operational Land Imager. *Remote Sensing*, *6*(10), 10232-10251. [doi:10.3390/rs61010232](http://www.mdpi.com/2072-4292/6/10/10232)

Brandt, J. (2008). Identifying wild rice beds in northern Wisconsin using multi-temporal Landsat imagery*.* Retrieved from <https://www.researchgate.net/publication/259739250_Identifying_Wild_Rice_Beds_in_northern_Wisconsin_using_multi-temporal_Landsat_imagery>

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5-32. doi:10.1023/A:1010933404324

Brummitt, N., & Bachman, S. (2010). Plants under pressure – a global assessment. The first report of the sampled Red List Index for plants. Royal Botanic Gardens, Kew, UK. Retrieved from <https://www.researchgate.net/publication/322682292_Plants_under_pressure_-_a_global_assessment_The_IUCN_Sampled_Red_List_Index_for_Plants>

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*.* https://doi.org/10.1016/j.rse.2017.06.031

Jin, S., & Sader, S. A. (2005). Comparison of time series tasseled cap wetness and the normalized difference moisture index in detecting forest disturbances. *Remote sensing of Environment*, *94*(3), 364-372. <https://doi.org/10.1016/j.rse.2004.10.012>

Kauth, R. J., & Thomas, G. S. (1976). The tasselled cap--a graphic description of the spectral-temporal development of agricultural crops as seen by Landsat. *LARS Symposia*. Retrieved from <https://docs.lib.purdue.edu/lars_symp/159/>

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*(4), 1496-1508. <https://doi.org/10.2135/cropsci2012.10.0585>

Meilleur, B. A., & Hodgkin, T. (2004). *In situ* conservation of crop wild relatives: status and trends. *Biodiversity & Conservation*, *13*(4), 663-684. <https://doi.org/10.1023/B:BIOC.0000011719.03230.17>

Minnesota Department of Natural Resources. (2008). *Natural wild rice in Minnesota: A wild rice document submitted to the Minnesota Legislature by the Minnesota Department of Natural Resources*. Minnesota Legislative Reference Library. Retrieved from https://www.leg.state.mn.us/docs/2008/mandated/080235.pdf

Myrbo, A., Swain, E. B., Engstrom, D. R., Coleman Wasik, J., Brenner, J., Dykhuizen Shore, M., ... Blaha, G. (2017). 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. https://doi.org/10.1002/2017JG003787

Nelson, A., Setiyono, T., Rala, A. B., Quicho, E. D., Raviz, J. V., Abonete, P. J., … Ninh, N. H. (2014). Towards an operational SAR-based rice monitoring system in Asia: Examples from 13 demonstration sites across Asia in the RIICE Project. *Remote Sensing*, *6*(11), 10773-10812. <https://doi.org/10.3390/rs61110773>

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

Stohlgren, T. J., Ma, P., Kumar, S., Rocca, M., Morisette, J. T., Jarnevich, C. S., & Benson, N. (2010). Ensemble habitat mapping of invasive plant species. *Risk Analysis: An International Journal*, *30*(2), 224-235. <https://doi.org/10.1111/j.1539-6924.2009.01343.x>

Suszkiw, J. (2014). National inventory takes stock of crops’ wild relatives*. United States Department of Agriculture, Agricultural Research Magazine*. Retrieved from <https://agresearchmag.ars.usda.gov/AR/archive/2014/Jan/crops0114.pdf>

United States Department of Agriculture Farm Service Agency. (2015, 2017). National Agriculture Imagery Program (NAIP) [Image collection], accessed 20 February 2019. Retrieved from <https://code.earthengine.google.com/>

United States Department of Agriculture Forest Service. (2014, October 8). USDA Forest Service and Agricultural Research Service crop wild relatives partnership projects: The importance of research natural areas, wilderness areas, and more. *USDA Forest Service Botany in the News.* Retrieved from <https://www.fs.fed.us/wildflowers/news/documents/BotanyInTheNews_CWR_RNAs_R9.pdf>

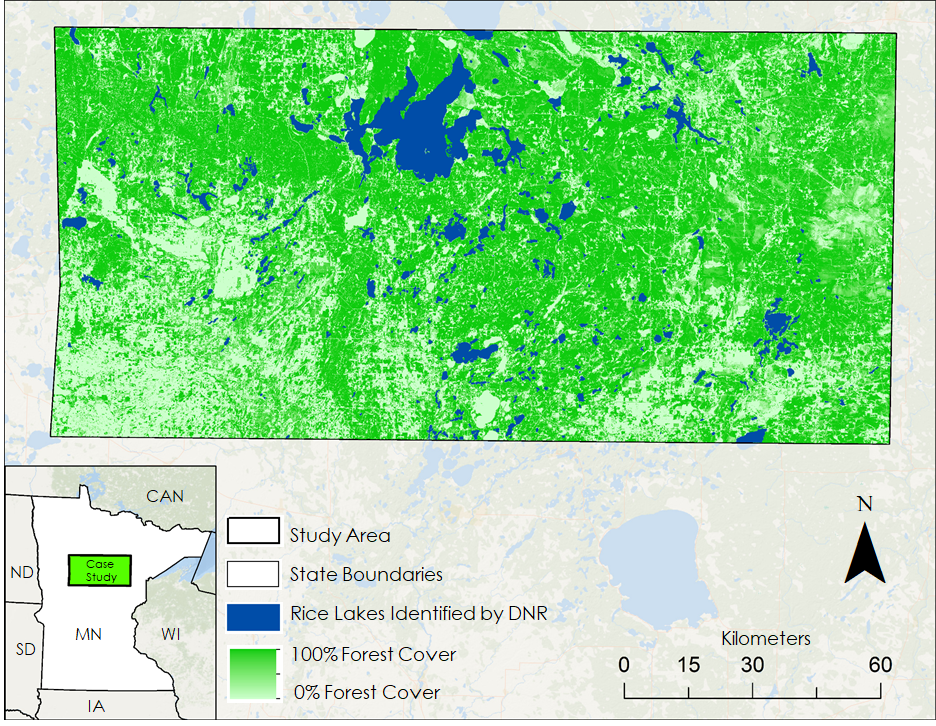
United States Geological Survey Earth Resources Observation and Science Center. (2014). Landsat 8 OLI Level-2 Surface Reflectance (SR) Science Product [Dataset], accessed 20 February 2019. <https://doi.org/10.5066/f78s4mzj>

Walker, K., Carver, D., LaRoe, J., & Whittemore, C. N. (2018). Minnesota & Texas Agriculture & Food Security: Employing NASA Earth Observations to model current and historic distribution of crop wild relatives, in support of USDA ARS genetic resource conservation efforts. *Technical Report.* *NASA DEVELOP National Program: Colorado ­– Fort Collins Node.*

Xiao, X., Boles, S., Liu, J., Zhuang, D., Frolking, S., Li, C., ... & Moore III, B. (2005). Mapping paddy rice agriculture in southern China using multi-temporal MODIS images. *Remote sensing of environment*, *95*(4), 480-492. https://doi.org/10.1016/j.rse.2004.12.009

# 9. Appendices

**Appendix A.** Sentinel-2 case study area.



**Appendix B.** Data acquisition and processing.

Table B1

*Spectral data utilized in satellite detection modeling of emergent aquatic vegetation*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Product | Spatial Resolution (meters) | Temporal Resolution | Source | Description |
| TOA Reflectance  (Bands 2-7) | 30 | 2015-2017 | Landsat 8 OLI TOA Reflectance Tier 1 | Reflectance measured at the top of the atmosphere, without atmospheric corrections applied |
| Spectral Indices  (NDVI, NDMI) | 30 | 2015-2017 | Landsat 8 OLI TOA Reflectance Tier 1 | Combinations of spectral reflectance from two or more wavelengths that indicate the relative abundance of features on the Earth’s surface |
| Tasseled Cap Components (Brightness, Greenness, Wetness) | 30 | 2015-2017 | Landsat 8 OLI TOA Reflectance Tier 1 | A linear transformation of spectral bands, used for vegetation mapping to distinguish bright, green, and wet surfaces, derived from Kauth-Thomas 1976 |
| Surface Reflectance | 10 - 20 | 2017 | Sentinel 2 MSI Level-1C | Reflectance measured at the Earth’s surface, with atmospheric corrections applied |
| Radar Backscatter | 10 | 2015-2017 | Sentinel 1 SAR C-Band Synthetic Aperture Radar | Wave scattering measured at the Earth’s surface used to understand changes in surface roughness |

Table B2

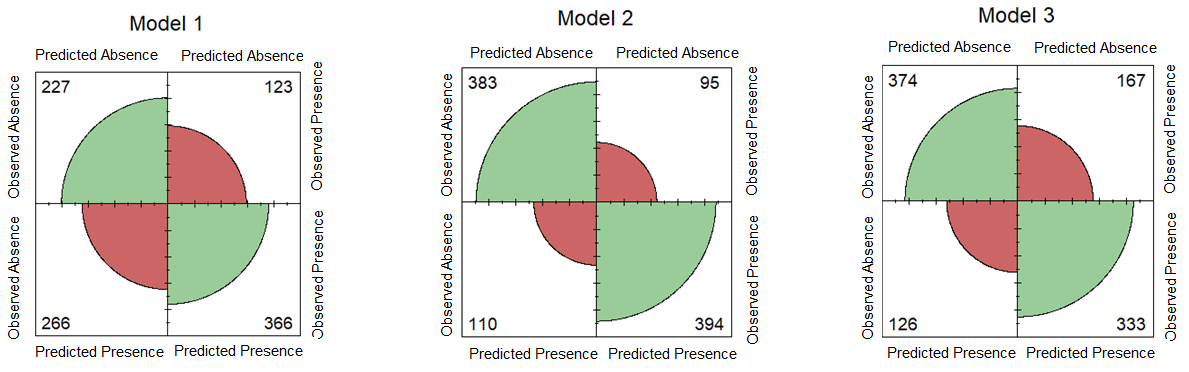
*Spectral and radar derivatives utilized in satellite detection modeling of emergent aquatic vegetation*

|  |  |  |  |
| --- | --- | --- | --- |
| Spectral Derivative | Equation | Description | Source |
| Normalized Difference Vegetation Index (NDVI) | NDVI = (NIR – Red) / (NIR + Red) | A remote sensing indicator that can be used to assess vegetation content | Xiao et al, 2005 |
| Normalized Difference Moisture Index (NDMI) | NDMI = (NIR – SWIR) / (NIR + SWIR) | A remote sensing indicator that can be used to assess water content | Jin & Sadar, 2005 |
| Tasseled Cap Wetness | Blue (.1511) + Green (.1973) + Red (.3283) + NIR (.3407) + SWIR (-.7117) + SWIR2 (-.4559) | A linear transformation of spectral bands, used to distinguish wet surfaces and measure water content | Kauth & Thomas 1976 |
| Tasseled Cap Greenness | Blue (-.2941) + Green (-.243) + Red (-.5424) + NIR (.7276) + SWIR (.0713) + SWIR2 (-.1608) | A linear transformation of spectral bands, used to distinguish vegetated surfaces and measure vegetation content | Kauth & Thomas 1976 |
| Tasseled Cap Brightness | Blue (.3029) + Green (.2786) + Red (.4733) + NIR (.5599) + SWIR (.508) + SWIR2 (.1872) | A linear transformation of spectral bands, used to distinguish bright surfaces and measure brightness content | Kauth & Thomas 1976 |
| Radar Range | Radar Median T1 – Radar Median T2 | The range of the radar backscatter signal over the growing season | Minnesota Agriculture, Spring 2019 |
| Radar Variance | Image result for sample variance formula | The variance of the radar backscatter signal over the growing season | Minnesota & Texas Agriculture, Spring 2018 |

**Appendix C**. Aerial comparison of wild rice and waterlilies over NAIP imagery.



**Appendix D.** Accuracy assessments.



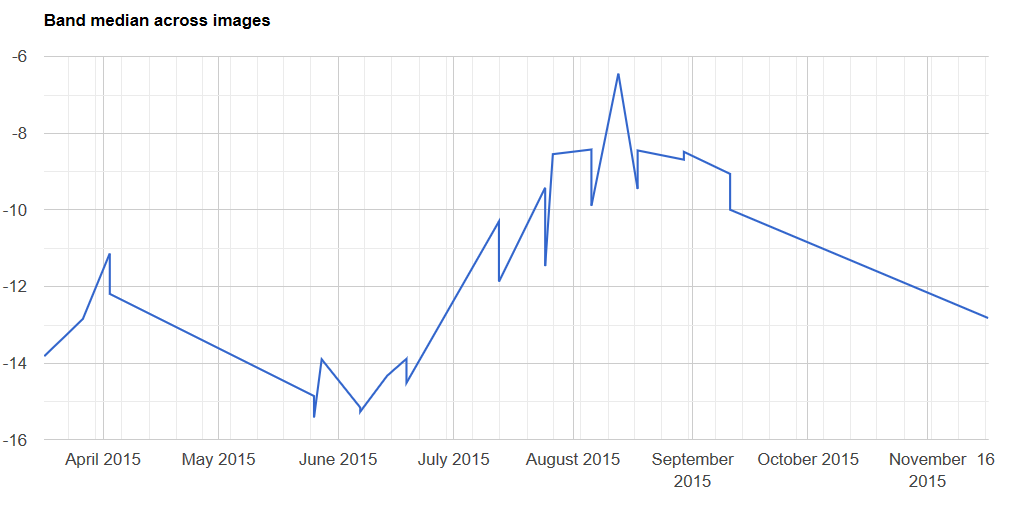
*Figure D1.* Validation confusion matrices

Table D1

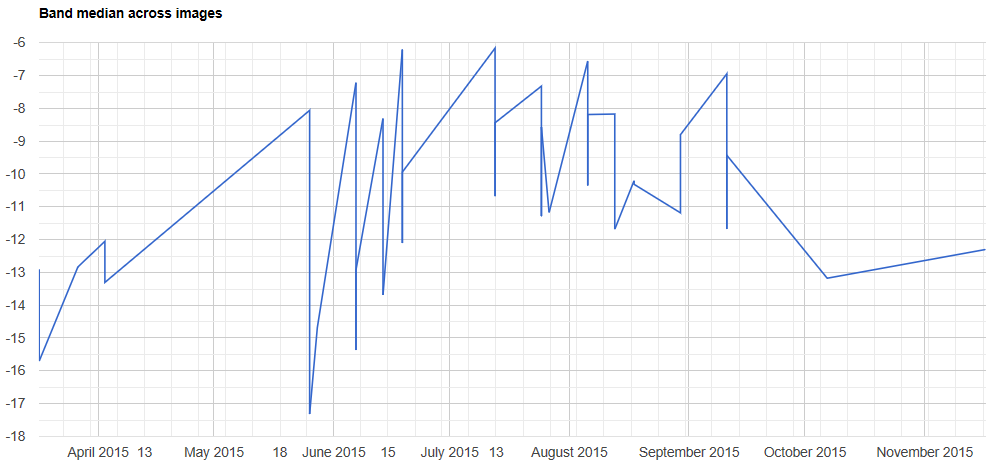
*Additional model evaluation metrics*

|  |  |  |  |
| --- | --- | --- | --- |
| Model # | AUC | Sensitivity | Specificity |
| 1 | 0.94 | 0.92 | 0.95 |
| 2 | 0.83 | 0.83 | 0.83 |
| 3 | 0.67 | 0.65 | 0.69 |
| 4 | 0.85 | 0.83 | 0.88 |

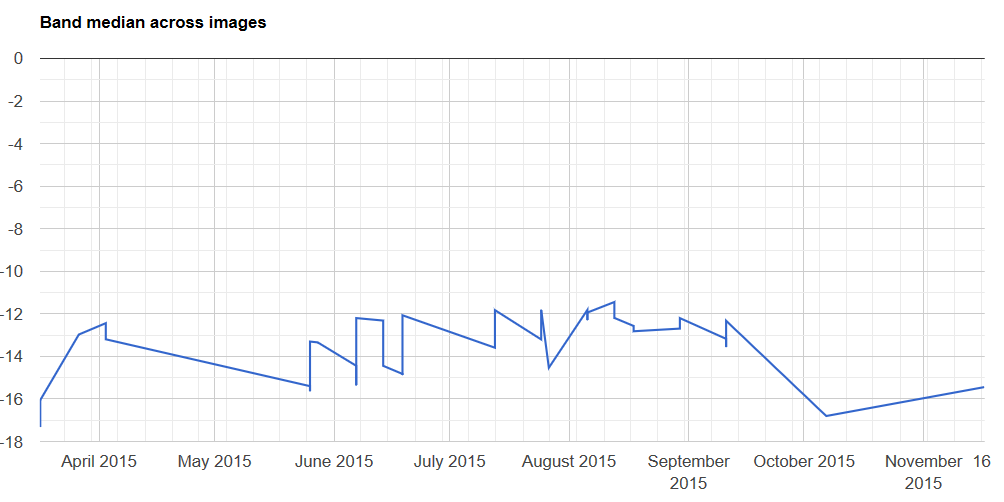
**Appendix E**. Radar signatures of wild rice, bulrushes, waterlilies, and cattails.

******

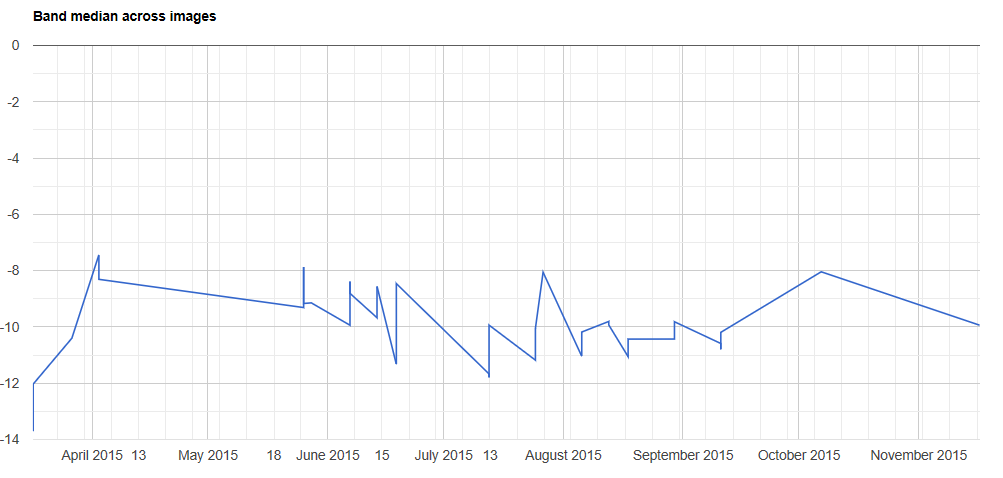
*Figure E1.* The radar signature of wild rice.

******

*Figure E2.* The radar signature of bulrushes.

******

*Figure E3.* The radar signature of waterlilies.

******

*Figure E4.* The radar signature of cattails.