**NASA DEVELOP National Program**

Colorado – Fort Collins

*Spring 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**

Final Draft – March 29th, 2018

Kaitlin Walker (Project Lead)

Daniel Carver

Jillian LaRoe

Charles Nicholas Whittemore

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

Nicholas Young, Colorado State University, Natural Resource Ecology Laboratory (Science Advisor)

Tony Vorster, Colorado State University, Natural Resource Ecology Laboratory (Science Advisor)

Brian Woodward, Colorado State University, Natural Resource Ecology Laboratory (Science Advisor)

# 1. Abstract

Northern wild rice (*Zizania palustris L.)* and Texas wild rice (*Zizania texana*) provide valuable ecosystem services, food sources, economic development, and cultural resources to local populations in Minnesota and Texas. Research on crop wild relatives, wild plants closely related to cultivated plants, is imperative to understanding gene flow and genetic diversity of harvested species. The United States Department of Agriculture (USDA) Agricultural Research Service (ARS) is responsible for conserving the genetic diversity of valuable species, such as wild rice. However, this organization lacks insight to the geographic distribution of *Zizania* populations. NASA Earth observations, including Landsat 5 Thematic Mapper, Landsat 8 Operational Land Imager and the Shuttle Radar Topography Mission version 3 were used to create models to detect wild rice presence. The team provided partners at the USDA ARS with distribution maps for northern wild rice and Texas wild rice populations in 2005 and 2015. Partners at USDA ARS will apply the end products to effectively enable strategic ecological planning, and better target field collections for species conservation.

**Keywords**

Crop wild relatives, Landsat 5, Landsat 8, Minnesota, random forest, Sentinel-1 SAR, Texas, USDA

# 2. Introduction

* 1. ***Background Information***

Domesticated crops are descendants of wild crop relatives. Over thousands of years of intensive breeding for specific traits, modern crop varieties are often dramatically different both phenologically and genetically compared to their wild ancestors. Native populations of wild crop relatives exist across the landscape and contain historical genetic traits that may no longer be present in modern crops due to breeding and cultivation practices. These traits found in crop wild relatives may be used to increase biophysical resiliency, diseases resistance, and genetic diversity in modern food staples (Seiler, Qi, & Marek, 2017). The Plant and Animal Genetic Resources Preservation (ARS) section of the United States Department of Agriculture’s (USDA) goal is to preserve this genetic diversity for the security of our food resources into the future (United States Department of Agriculture Agricultural Research Service). A challenge facing the ARS is understanding the extent and range of specific wild crop relatives throughout the United States. This study employs NASA Earth Observations to aid the USDA in determining the extent of wild rice (*Zizania*) to ensure the ARS can continue to collect and preserve these iconic North American species.

Wild rice has been harvested in the Great Lakes region for two thousand years (Norrgard, 2008). The Ojibwe tribe of the northern Great Lakes region have a strong cultural connection to wild rice, which is referred to as “manoomin.” Their history states that their migration from the East Coast was guided by an attempt to find “food on water”. It is believed that this close relationship between people and wild rice has influenced the current distribution of rice throughout the northern Great Lakes region (Norrgard, 2008). The cultural tradition of harvesting wild rice is continued in current times. Individuals gently knock the rice grains from the stands of rice into a canoe with a pole specifically made for harvesting. In modern times, the collection of wild rice is an important part of the cultural identity of indigenous peoples and also represents a source of economic development for the region (Norrgard, 2008). The first study that attempted to use remotely sensed imagery to detect wild rice distribution was done on behalf of the Ojibwe tribe (Frohn and Price, 2003) to estimate the lost wild rice economic production due to heavy storms early in the plants growth cycle.

There are three species of North American wild rice: Northern wild rice (*Zizania palustris* L.) is found across Canada and the conterminous United States; annual wild rice (*Zizania aquatica*) is found across Canada and the eastern United States; and Texas wild rice (*Zizania texana*) is endemic to the San Marcos River in Hays County, Texas. *Zizania texana* is an endangered species, and only grows along a short stretch of the San Marcos River in Hays County, Texas (Breslin, S. L. 1997). This stretch of river is located within a populated area which makes conservation urgent. Our methods focused on *Zizania palustris* L. in Minnesota due to the widespread extent of wild rice. After refining the methodology for *Zizania palustris* L. we modelled *Zizania texana* in Hays County, Texas.

Wild rice is an aquatic plant which grows in monotypic stands in fresh water bodies, wetlands, and along the stream channel (Price, 2012). Stands of wild rice can grow up to 100 hectares in area (Price, 2012). Wild rice can be found growing in water depths between 0 and 1.5 meters. As a result, it is highly susceptible to changes in water level (Norrgard, 2008). It is a perennial plant that is known to occupy a single location for many years, though the density and size of the stand can change significantly between years (Price, 2012). This plant has a unique seasonal phenology that includes seven distinct stages. We summarized these stages into four distinct periods for remote sensing purposes. In the first open water stage (December to late May), wild rice is submerged beneath the water surface. During the second emergent stage (June to July), the rice grows horizontally along the water surface until it is rigid enough to emerge from water and stand perpendicularly. At the third peak growth stage, the plant stands and matures to produce seeds (August to early September). During this stage, the plant can be up to 1 meter in height (Norrgard, 2008). The peak growing phase corresponds with the traditional harvest time (Norrgard, 2008). During the final senescence stage (October - November), rice stands lose their green color and eventually fall into the water. Our team’s understanding of the phenological development of wild rice allowed us to identify specific time period and utilize imagery representative of these time periods to differentiate this species from other aquatic vegetation.

Previous studies have successfully modelled rice species using either Landsat 5 Thematic Mapper (TM), Landsat 8 Operational Land Imager (OLI), or Sentinel-1 C-band Synthetic Aperture Radar (SAR) remotely sensed data (Brandt, 2008). Brandt (2008) used Landsat imagery to define multiple land use and land change classes to incorporate continuous cover. SAR has proved useful in identifying the unique spectral signature of wild rice due to the sensor’s ability to transmit through cloud cover and the sensors high temporal resolution. Nelson (2014) found SAR more effective at monitoring rice over time in Asia than relying on spectral imagery alone. Masking layers, such as bathymetry and hydrology, have also been used as a means to define suitable growing habitat for wild rice (Anderson et al., 2011; Brandt, 2008)

The goals of this study were to develop an accurate, scalable, and reproducible methodology to detect the presence of wild rice across the landscape. This methodology will enable our partners at the USDA ARS division to more effectively incorporate NASA Earth Observations into their mission.

* 1. ***Project Partners & Objectives***

DEVELOP partnered with the USDA Genetics and Seed Lab in Fort Collins, CO. The USDA is interested in preserving the genetic diversity within crop wild relative species. The results of our project will help the USDA make conservation decisions for *Zizania palustris* L., and *Zizania texana*. The USDA will use the models and methodology created from our research to locate crop wild relatives for preservation. The outputs from our models will assist field scientists in locating geographically isolated areas within each species range and help guide decisions about the collection of samples.

Our study objectives include: 1) testing the feasibility of using NASA Earth observations to detect crop wild relatives, 2) creating a presence map of populations of *Zizania palustris* L. in Minnesota and *Zizania texana* in Hays County, Texas, 3) tracking the distribution of wild rice from 2005 to 2015, and 4) creating a toolset and tutorial for use in other regions.

# 3. Methodology

***3.1 Data Acquisition***

All imagery was acquired and processed using Google Earth Engine. Spectral imagery processed to Tier 1 surface reflectance from Landsat 5 TM and Landsat 8 OLI were used. A 25% threshold for cloud cover was used to limit the effect of clouds. Sentinel-1 extra wide swath C-band SAR imagery was geometrically corrected and preprocessed. This preprocessing was completed by the creators of Google Earth Engine and provided through the Google Earth Engine interface. We obtained a corrective algorithm from the Google Earth Engine Developer Forum to reduce edge effects when creating mosaicked SAR images.

Table 1

*Data Products and Acquisition*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Platform and Sensor** | **Data Product** | **Imagery Dates** | **Bands Utilized** | **Resolution** |
| Landsat 5 TM | USGS Landsat 5 TM Surface Reflectance (Orthorectified) | June - October, 2005 | B1, B2, B3, B4, B5, B7 | 30 meters |
| Landsat 8 OLI | USGS Landsat 8 Surface Reflectance (Orthorectified) | June - October, 2015 | B1 - B7 | 30 meters |
| Sentinel-1 SAR GRD | Sentinel-1 SAR GRD: Synthetic Aperture Radar Ground Range Detected, Level-1 | June - October, 2015 | C-band: Vertical-Vertical (VV) polarization  Extra-wide swath capture | 10 meters |
| National Agricultural Imagery Program | Digital Ortho Quarter Quad Tiles (DOQQs) | June - August, 2005  August - October, 2015 | Blue, Green, Red  NIR(Near-Infrared only available for 2015) | 1 meter |

***3.2 Data Processing***

We visualized high resolution aerial imagery from the National Agricultural Imagery Program (NAIP) for the 2015 growing season using the Google Earth Engine (GEE) interface. These images were captured during the months of September and October, and all tiles were displayed across the study area for opportunistic sampling. We used a false color composite (NIR, B, G) to aid in distinguishing wild rice from other vegetation for ocular sampling.

Ancillary datasets were prepared as validation tools. We utilized a bathymetry layer available from the Minnesota Department of Natural Resources Geospatial Information Office to depths corresponding to known wild rice habitat (Norrgard, 2008). The Minnesota Department of Natural Resources provides a statewide inventory of lakes containing rice and estimated percent rice cover, based on historical field surveys (Drotts, 1980). Sample points were dropped within lakes identified in the DNR statewide inventory, using the bathymetry data as a guide.

***3.3 Data Analysis***

***3.3.1 Sample Generation***

Based on the limited availability of field datasets for our study area and date range, we generated additional opportunistically sampled training points. These points were ocularly sampled using 2015 NAIP imagery at 1 meter resolution in GEE (United States Department of Agriculture Farm Service Agency). Field data provided by our partners was used as reference presence points for the identification of wild rice stands in Minnesota; this reference field data could not be used for modelling because it was not collected for remote sensing purposes. We generated evenly distributed points across three Landsat scenes that contain many lakes with wild rice in the state of Minnesota. A Landsat 30m grid was overlaid onto the NAIP imagery to ensure that the opportunistically sampled locations were representative of a single Landsat cell. Presence points were defined as locations where a Landsat cell area was estimated to contain at >50% wild rice cover. Absence points were selected at locations that contained no wild rice. Absence points 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, with presence and absence classes approximately equally represented. A subset consisting of 310 of these points were used for single-scene modelling. These presence and absence data points were used as inputs for modelling efforts.

***3.3.2 Variable Selection***

To distinguish wild rice from other aquatic vegetation, all spectral bands and a series of indices from Landsat 8 Surface Reflectance and C-band SAR images from Sentinel-1 were loaded into Google Earth Engine in three specific time frames representing the open water, emergent, peak growth, and senescence stages of wild rice. Landsat images containing less than 20% cloud cover were selected. Sample locations of wild rice presence and absence were overlaid onto these images and the values representing a specific cell location were extracted to a corresponding point. These values were exported from Google Earth Engine and were used as the input for the variable selection process in R (version 3.4.4).

The Variable Selection using Random Forest (VSURF, version 1.0.3) library in RStudio (version 3.4.4) was used to determine which spectral indices were most effective at distinguishing wild rice from the surrounding landscape and other aquatic vegetation. Predictors were ranked in order of importance and displayed in a correlation matrix. Predictors containing a Pearson’s correlation coefficient >= |0.75| were removed based on which predictor had the greatest significance to model performance. The final series of predictors and their associated values were exported as a csv.

Table 1. Random forest model regions and temporal scales.

|  |  |
| --- | --- |
| **Region** | **Year** |
| Minnesota State | 2005 |
| Minnesota State | 2015 |
| Minnesota, Path/Row 28/28 | 2015 |
| Hays County, TX | 2016 |

***3.3.3 Modeling a Single Scene in Minnesota***

A subset of sampling points that fell with the Landsat scene Path/Row 28/28 was applied to our variable selection methodology. Based on the predictors defined by the VSURF (version 1.0.3) process the associated imagery was called into Google Earth Engine. Using the csv containing the reflectance values of the selected predictors, a random forest model was trained and used to predict the location of wild rice across the Landsat scene. This binary image was exported as a GeoTIFF for the area within our study region (Figures 3 & 4).

***3.3.4 Scaling Analysis: the State of Minnesota***

To predict wild rice across the state of Minnesota, the methodology was adapted to account for the challenges of cloud cover. No single set of images could provide cloud cover less than 40% for the entire state. Three image collections of Landsat scenes were created; June-July, July-August, and August-September. A median function was applied to each collection resulting in three composite images with less than 20% cloud cover across the entire state. This process reduced any potential impacts of clouds within the composites. Once generated, the same methodology of extracting values, determining variable importance, predicting locations of wild rice, and masking were completed for these median composites as described in the Variable Selection methodology. A wild rice presence map for Minnesota was produced from this process (Figures 5 & 6).

***3.3.5 Applying Models to Hays County, Texas***

The extent of wild rice within Texas limited our ability to ocularly sample for presence and absence of wild rice. Under the assumption that the spectral and SAR indices used to detect wild rice in Minnesota would also represent wild rice in Texas, a random forest model was trained using the 2015 data for Minnesota. The model was applied to the equivalent imagery in Texas. The binary map produce was masked to the San Marcos River and the resulting image is the final presence map for wild rice in Hays County, Texas (Figures 9 & 10).

***3.3.6 Historical Distribution modelling***

The spatial extent of wild rice varies annually based on environmental conditions. Our sampling methodology was based on 2015 imagery. To apply our methodology to future and past years, a filtering process was developed to subset our 2015 sampling points based on the likelihood they represent wild rice in different time frames. Processing steps are as follows:

1. The filtering process utilized the +/- standard reflectance of extracted values for 2015 (Figure 1, box 1).
2. Imagery from 2005 was used with the same extracted geographic points.
3. Subset of 2005 presence locations with a high confidence of similar reflectance properties to that of 2015.

An R script was created that imported the 2015 reflectance values for all predictor variables. The data was subsetted based on presence and absence and the mean and standard deviation of each predictor variable was calculated. An acceptable range of variability was defined by +/- one standard deviation from the mean. The high and low value for presence and absence points for each predictor was recorded in a data frame.

The data containing reflectance values from the 2005 imagery were read into the R script and subset into presence and absence. The reflectance value for each location was tested to see if it was within the acceptable range defined by +/- one standard deviation of the mean of the 2015 reflectance values. To account for the variance in bit depth between the sensors, a ratio of 0.0625 was applied to the mean and standard deviation of all individual bands from the 2015 predictor set. The binary results were recorded for every sample location for all predictors and a sum of these values was determined. Locations where ≥ 50% of predictors remained within the acceptable range were assumed to represent wild rice presence. All values that did not meet this condition were removed from the dataset. The remaining values were brought into GEE where they were used to train a random forest model for the 2005 time period.

Acceptable Range +/- 1 std from mean

Calculate Mean and Standard Deviation

Known Rice Location Reflectance Value

Test if value falls within acceptable range

Import point values from other year

Apply binary true false based on test result

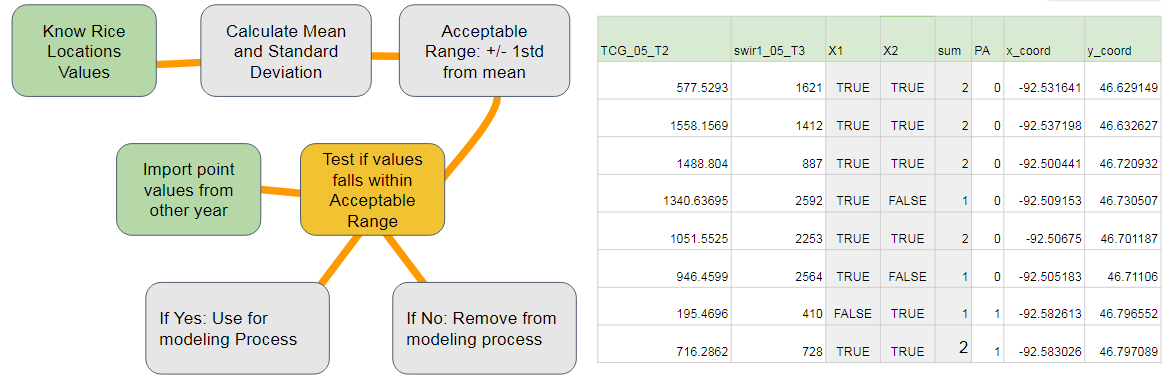


Figure 1: Historical modelling by Filtering Flowchart

***3.3.7 Masking and Differencing***

To identify emergent vegetation, variance was calculated for each cell of a time series of Sentinel-1 SAR Vertical Vertical polarized images captured from May through October using Google Earth Engine. Multiple threshold values were applied to the corresponding image to define the range of values characteristic of emergent vegetation. A binary map of presence of emergence vegetation was produced based on the selected threshold and exported with a 30m cell size within our study region.

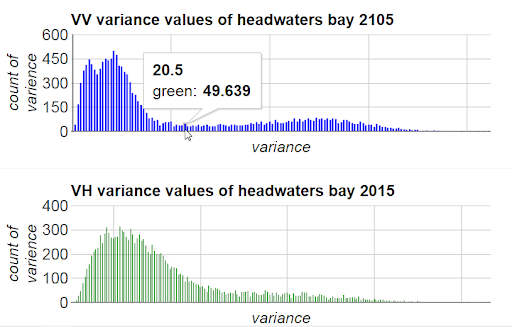


Figure 2. VV Variance Values of Headwaters Bay (2015).

The two binary classifications, emergent vegetation and wild rice presence, were brought into ArcMap and multiplied producing a final image in which presence was determined by areas where both images were defined as presence. This process was not completed for the 2005 model due to the lack of SAR imagery for that time period.

All model outputs, with the exception of Hays County, Texas, were masked to three classes in the National Land Cover Dataset (Home *et al*., 2015). Selected cover classes were: emergent herbaceous wetlands, woody wetlands, and open water. These classes were used to ensure that the final product only predicted wild rice in ecological conditions where the species is known to grow: open water, wetlands, and riparian zones (Norrgard, 2008). End products were binary maps of wild rice for each of the four models: 2005 state of Minnesota, 2015 state of Minnesota, 2015 Landsat Path/Row 28/28, and 2016 Hays County, Texas.

# 4. Results & Discussion

***4.1 Results***

Table 2. Selected model evaluation statistics.

|  |  |  |
| --- | --- | --- |
| Model | Out of Bag Error (%) | Area Under Receiver Operator Curve (AUC) |
| Single Scene (2015) | 7.1 | 0.697 |
| Full State Minnesota (2015) | 8.54 | 0.6838606 |
| Full State Minnesota (2005) | 17.01 | 0.674117 |

Table 3. Top predictors by model.

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Landsat Path/Row 28/28 | Full State Minnesota | Full State Minnesota |
| Year | 2015 | 2015 | 2005 |
| OOB | 7.10% | 8.54% | 17.01% |
| Top Predictors | Ultra Blue 2015 T2 | Normalized Difference Water Index 2015 T2 | Short-wave infrared 1 2005 T3 |
|  | Modified Normalized Difference Water Index 2015 T1 | Short-Wave Infrared 2015 T1 | Tasseled Cap Greenness 2005 T2 |
|  | Normalized Difference Vegetation Index 2015 T3 | Radar: June - August |  |
|  | Short-Wave Infrared 2015 T2 | Near-Infrared 2015 T3 |  |
|  | Radar: August - October | Normalized Difference Vegetation Index 2015 T3 |  |
|  | Red 2015 T3 | Tasseled Cap Wetness 2015 T3 |  |
|  |  | Green 2015 T3 |  |
| Tasseled Cap Coefficients for Landsat 8 (Baig, Zhang, Shuai, & Tong, 2014) and Landsat 5 (Huang, Wylie, Yang, & Homer, 2002). | | | |

Table 3 Legend

|  |
| --- |
| T1 = Spectral Predictors in June/July |
| T2 = Spectral Predictors in August |
| T3 = Spectral Predictors in September |

***4.1.1 Modelling a Single Scene in Minnesota.***

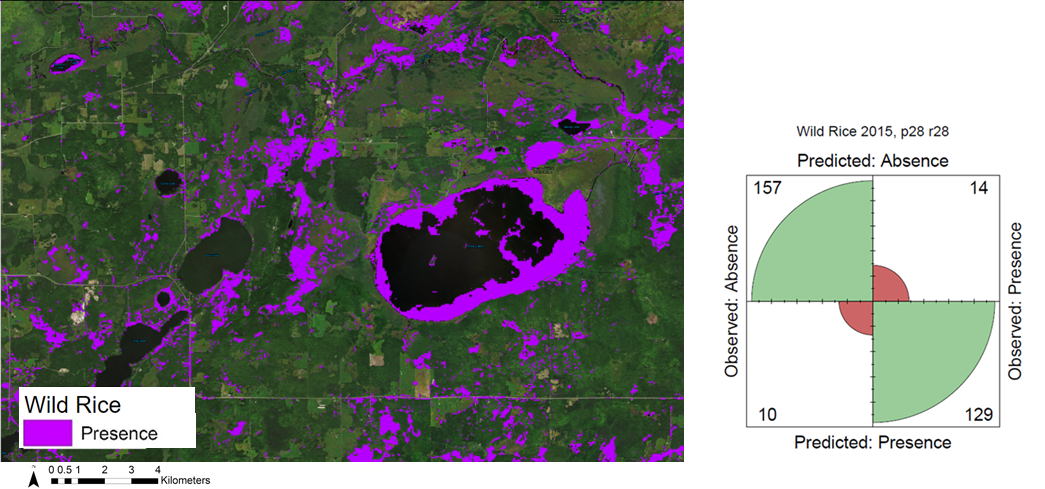
******

Figure 3. Landsat Path/Row 28/28 Model (2015).

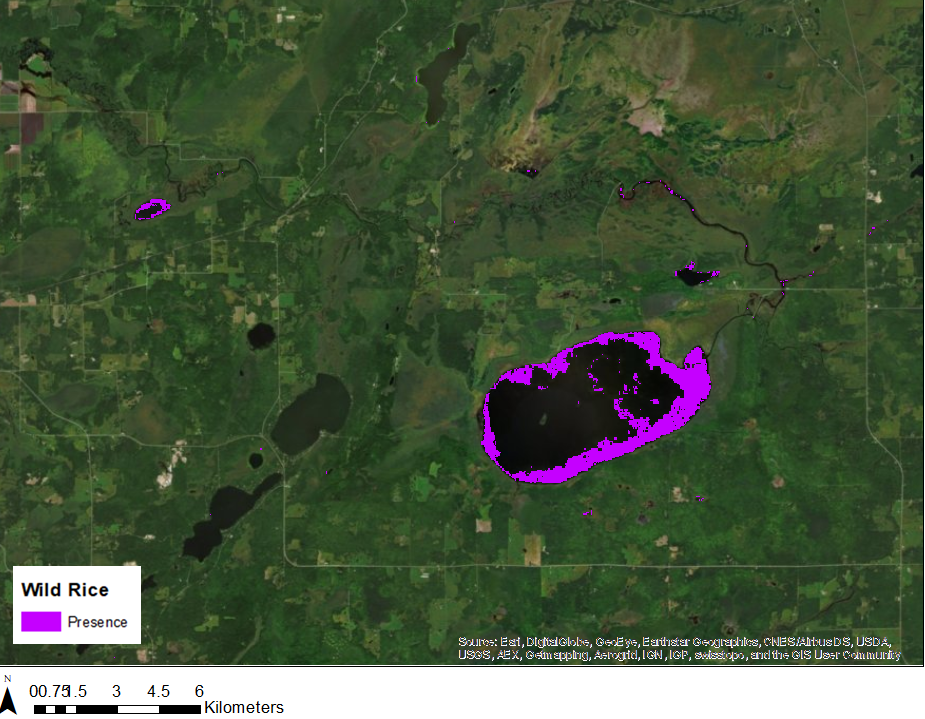


Figure 4: Landsat Path/Row 28/28 Model (2015) Masked by SAR and NLCD.

***4.1.2 Scaling Models to Minnesota State.***

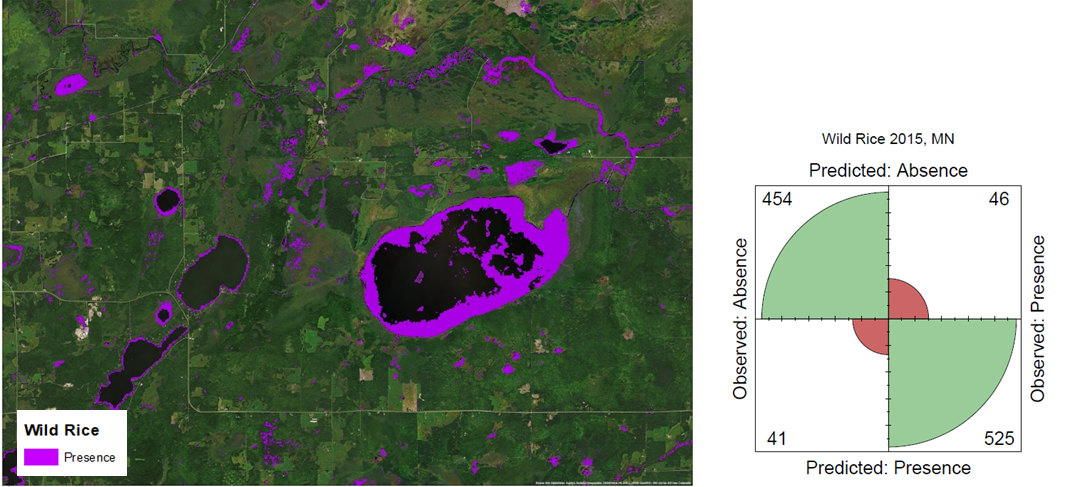
******

Figure 5. Minnesota State Model (2015).

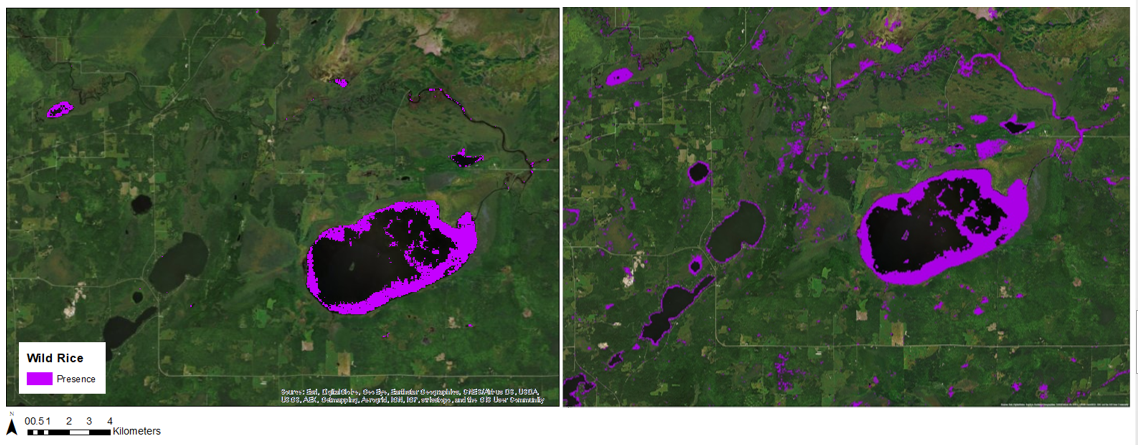
******

Figure 6. Minnesota State Model Masked by SAR and NLCD (2015).

***4.1.3 Applying models over time.***

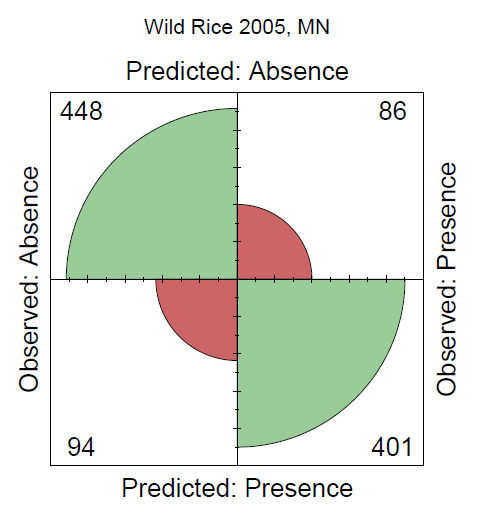
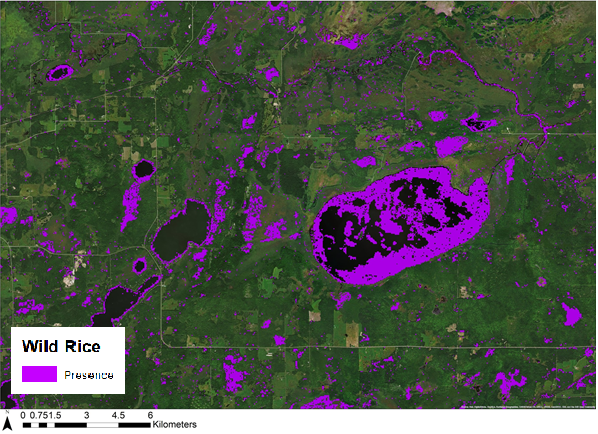
******

Figure 7. Minnesota State Model Unmasked (2005)

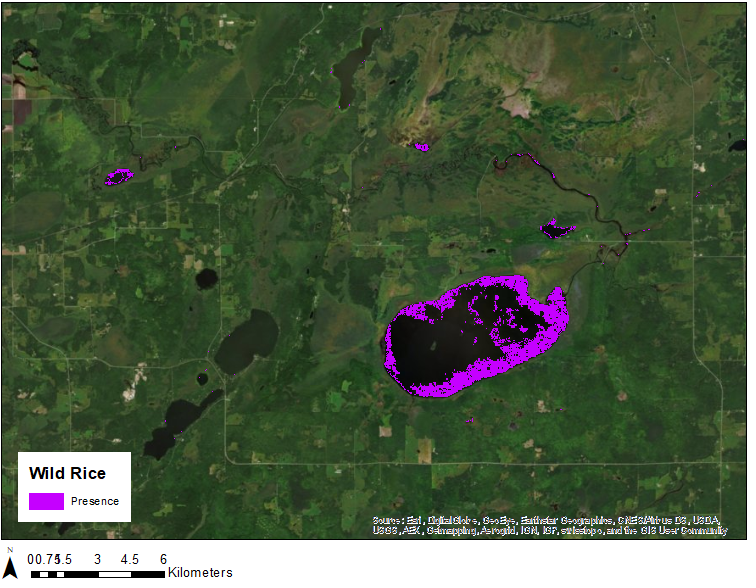
******

Figure 8. Minnesota State Model Masked by NLCD (2005)

***4.1.4 Applying models over geographic space.***

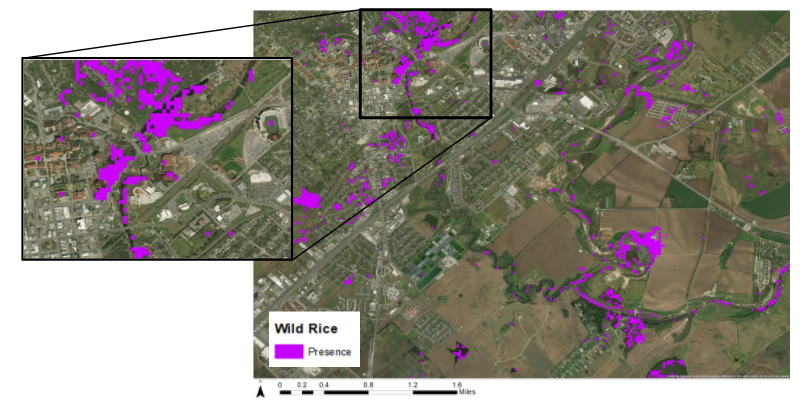
******

Figure 9. San Marcos River, Texas Model (2016).

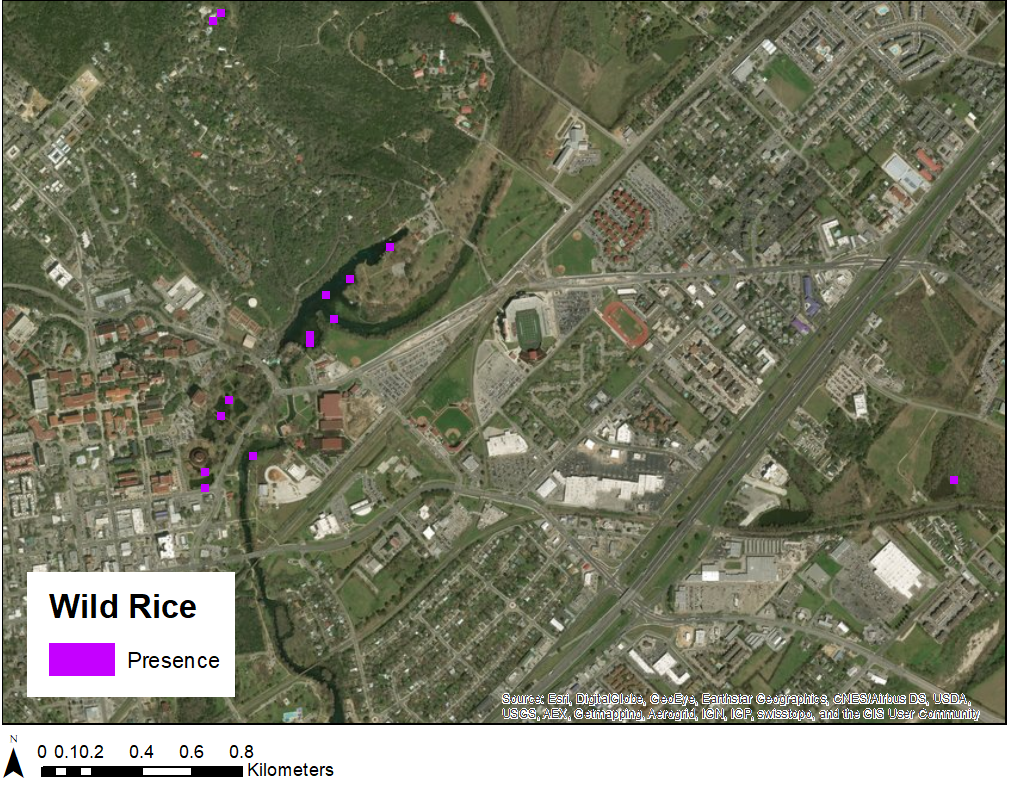
******

Figure 10. San Marcos River, Texas Model, Masked by SAR (2016).

***4.2 Discussion***

***4.2.1 Comparison of Minnesota Models***

The unmasked models generally predicted wild rice in some agricultural areas based on visual assessment. It is important to note that masking areas unsuitable for wild rice growth using the selected NLCD cover classes was crucial to minimize the exterior noise. Similarly, SAR C-band masking also eliminated areas easily identified by ocular inspection as other land cover types (Figures 4, 6, and 8).

The single scene model predicted more wild rice in areas outside of lakes than the Minnesota state models did, while the Minnesota state models predicted more wild rice within water bodies than the single scene model. The lower quantity of training points used for the single scene model may have resulted in more agricultural or urban land outside of lakes categorized as presence; we used 310 points for 2015, whereas we used 1,125 points for the Minnesota state model. Upon qualitative visual assessment, the full state of Minnesota model for 2015 appears to more accurately map wild rice in comparison to the Path/Row 28/28 model for 2015, although the OOB error is higher by less than two percent (Table 2). The single scene model over predicted wild rice presence within unrelated land cover types. Additional training points or ground verification may improve model accuracy and better refine this process, especially for the single scene model. The same auxiliary datasets used with the single scene model above were used again to mask the model outputs.

***4.2.2 Applying Models to Texas***

The team was able to apply the Minnesota state model to Hays County, Texas, but was unable to perform a quantitative assessment of model accuracy. We did not produce an independent presence/absence dataset for this study area due to the extremely limited geographic extent of *Zizania texana*. Upon visual inspection, the model again over predicted presence outside of water bodies or wetlands, and misclassified some small urban patches and surrounding fields as wild rice. Following masking with NLCD data, a fraction of the modeled presence area remained. The remaining areas were located in Spring Lake near the town of San Marcos, and a few single-cell patches along the San Marcos River.

***4.2.3 Historical modelling***

Model accuracy for the full extent of Minnesota in 2005 may have been lower than that of 2015 due to the lower NAIP image quality for this year. The top predictors for this model were selected independently of the other models using the same framework. The lack of radar availability for 2005 may have further decreased the accuracy of these results (Figure 7). The team explored an alternate methodology for predicting historical presence based on a filtering method to subset the sampling points for this reason.

***4.2.4 Future Work***

The methodology developed this term may serve as a strong foundation for using SAR and similar modelling processes to identify other crop wild relatives of interest, particularly aquatic species. Utilizing both SAR and Landsat imagery can highlight key traits and seasonal growth in aquatic vegetation. Future studies may increase accuracy by using a similar modelling approach in tandem with habitat suitability models for further refinement of distribution models. Initial exploration of an alternate methodology for predicting historical presence using the filtering method suggested that this methodology may be successfully adapted for future historical modelling efforts.Further exploration of these methods may help predict the distribution of crop wild relatives over additional time periods. To apply this methodology with other species, the ecological parameters in relation to spatial and temporal parameters will need to be catered to the species of interest to highlight its’ unique phenology.

This study may also be reproduced to continually monitor future changes in distribution and seasonality of wild rice in response to environmental and anthropogenic disturbances. Furthermore, our modelling process may be improved in future studies by eliminating areas mapped as presence that have a spectral reflectance outside the range of values of ocularly sampled training data.

# 5. Uncertainty and Error

***5.1 Wild Rice Growth Patterns***

Remote sensing of wild rice may be complicated by its cyclic nature. Since wild rice regularly cycles through three to five year oscillations in productivity, mapping the plant using remote sensing techniques may involve some inherent uncertainty (Norrgard, 2008). Mixed aquatic vegetation also presents a problem for identifying wild rice. Wild rice does not always grow in monotypic stands, so Landsat cells may contain mixed vegetation, potentially decreasing the accuracy of a model (Brandt, 2008).

***5.2 NAIP Sampling***

NAIP imagery from 2005 does not include the near infrared (NIR) band, which is necessary to create a false color image display to visually distinguish wild rice from other aquatic vegetation. Lower image quality resulted in a lower level of classification certainty when creating our ocularly sampled training data for 2005. Clear distinctions of wild rice and aquatic vegetation are not feasible without the NIR band. A filtering methodology was applied when modelling historical distribution to account for the inherent uncertainty of modelling different periods of time.

Our sampling process may be improved by using *in situ* field data collected specifically for remote sensing purposes. The team could not obtain field or expert validation of our ocularly sampled points, so some ocularly sampled points may have contained mixed or other aquatic vegetation. In this case, our maps may include the distribution of other species of aquatic vegetation. The team used SAR imagery masking to help combat this issue. Potential discrepancies between each individual sampler’s assessments may have introduced additional error.

# 6. Conclusions

This study examined the feasibility of three research objectives: 1) utilizing NASA Earth Observations to detect crop wild relatives, 2) creating a binary presence/absence map of populations of *Zizania palustris* L. in Minnesota and *Zizania texana* in Texas, and 3) modelling the distribution of wild rice in 2005 and 2015. These efforts have provided insights into using Earth observations to detect crop wild relatives, and how our partners may utilize our methods in the future:

1. The team successfully modeled wild rice presence in Minnesota and Texas using Landsat 5, Landsat 8, and Sentinel-1 SAR. We incorporated both spectral and SAR data in our models, and obtained higher qualitative accuracy upon visual inspection after applying SAR and NLCD masks.
2. We modeled wild rice presence in a larger geographic region and in another geographic area, the state of Minnesota and the San Marcos River in Texas respectively. Our model for the state of Minnesota in 2015 performed better than our model for a single scene in Minnesota. The 2016 Texas model was able to predict wild rice presence along areas of the river previously recorded as containing wild rice.
3. We investigated and established a methodology for predicting historic presence of crop wild relatives using a filtering method.
4. The use of Sentinel-1 SAR as a predictor improved model accuracy and the application of SAR variance mask improved the spatial accuracy of the model results. Thus, we recommend the use of SAR in conjunction with spectral imagery in future studies to detect crop wild relatives.

In summary, our partners at USDA can use our results and methodologies to inform future analyses of wild rice locations and spatial distributions. Our products will also aid in mapping other crop wild relatives using remotely sensed data.

# 6. Acknowledgments

Dr. Paul Evangelista (Colorado State University, Natural Resource Ecology Laboratory)   
Dr. Amanda West (Colorado State University, Natural Resource Ecology Laboratory)

Dr. Catherine Jarnevich (US Geological Survey)  
Nicholas Young (Colorado State University, Natural Resource Ecology Laboratory)  
Tony Vorster (Colorado State University, Natural Resource Ecology Laboratory)  
Brian Woodward (Colorado State University, Natural Resource Ecology Laboratory)

Dr. Colin Khoury (US Department of Agriculture, Agricultural Research Service)

This material contains modified Copernicus Sentinel data (2016), processed by ESA.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) 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 and cooperative agreement NNX14AB60A.

# 7. Glossary

**API** – Application programming interface

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

**ESA** – European Space Agency

**ESRI ArcGIS** – Environmental Systems Research Institute platform for image processing, data analysis, map creation, and end product generation

**GEE** – Google Earth Engine API, free software used for large scale image analysis and data acquisition

**Landsat** – A series of Earth-observing satellites which offers the longest continuous space-based record of Earth’s land cover

**NAIP** – National Agricultural Imagery Product, one meter resolution aerial imagery collected during the agricultural growing season in the continental United States

**NDMI** – Normalized Difference Moisture Index

**NDVI** – Normalized Difference Vegetation Index

**NIR** – Near infrared

**NLCD** – National Land Cover Database, a comprehensive land cover product from the Multiresolution Land Characteristics Consortium derived from decadal Landsat imagery and other supplemental datasets

**NNI** – Narrow near infrared

**OLI** – Operational Land Imager

**Quadrat** – A square plot used to isolate a standard unit of area for study

**R** – Open source software language and environment for statistical modelling, used for index calculations

**Random forest** – A classification modelling method trained using decision trees to guide analysis and prediction patterns of a large dataset

**SAR GRD** – Synthetic Aperture Radar, Ground Range Detected

**Sentinel-1** – A two-satellite constellation with the objective of Land and Ocean monitoring from the European Space Agency

**SRTM** – Shuttle Radar Topography Mission

**Surface reflectance** – The fraction of incoming solar radiation which is reflected from the surface of the Earth

**SWIR** – Short wave infrared spectral range

**TCB** – Tasseled cap brightness

**TCG** – Tasseled cap greenness

**TCW** – Tasseled cap wetness

**TM** – Thematic Mapper from the Landsat series which provided seven bands of image data (three in visible wavelengths, and four in infrared) at 30 meter resolution

**USDA ARS NPGS** – United States Department of Agriculture, Agricultural Research Service, National Plant Germplasm System

**VNIR** – Visible near infrared

# 8. References

Anderson *et al.* (2011). *Comparison of Wild Rice and Waterfowl Surveys*. Minnesota Department of Natural Resources.

Baig, M.H.A., Zhang, L., Shuai, T., and Tong, Q. (2014). Derivation of tasseled cap transformation based on Landsat 8 at-satellite reflectance. *Remote Sensing Letters 5*(5), 423-431.

Brandt, J. (2008). Identifying wild rice beds in northern Wisconsin using multi-temporal Landsat imagery. *Circle of Flight,* 1-30.

Breslin, S. L. (1997). *The impact of recreation on Texas wild-rice* (Unpublished doctoral dissertation). Southwest Texas State University, San Marcos, TX.

Drotts, G. (1980). *Wild Rice Distribution and Abundance in Minnesota. Minnesota Department of Natural Resources*. Retrieved from [http://files.dnr.state.mn.us/fish\_wildlife/wildlife/shallowlakes/statewide-inventory-wild-rice-water pdf](http://files.dnr.state.mn.us/fish_wildlife/wildlife/shallowlakes/statewide-inventory-wild-rice-waters.pdf)

ESA. (2017). Copernicus and Open Access Hub. European Commission: European Space Agency, 2014-2017. [Data set]. Retrieved from <http://scihub.copernicus.eu/>

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 the Environment, 202,* 18-27. https://doi.org/10.1016/j.rse.2017.06.031

Huang, C., Wylie, B., Yang, L., Homer, C., and Zylstra, Z. (2002). Derivation of a tasseled cap transformation based on Landsat 7 at-satellite reflectance. *International Journal of Remote Sensing, 23*(8), 1741-1748.

Nelson, A. (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.

Norrgard, R. (2008). *Natural Wild Rice in Minnesota* (Rep.). Retrieved from <http://files.dnr.state.mn.us/fish_wildlife/wildlife/shallowlakes/natural-wild-rice-in-minnesota.pdf>

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

Seiler, G., Qi, L., Marek, L (2017) Utilization of Sunflower Crop Wild Relatives for Cultivated Sunflower Improvement. *Crop Science, 57,* 1083–1101. <http://doi.org/10.2135/cropsci2016.10.0856>.

United States Department of Agriculture Agricultural Research Service. *About ARS: USDA ARS (2017*). Retrieved from <https://www.ars.usda.gov/about-ars/>

United States Department of Agriculture Farm Service Agency, *National Agriculture Imagery* Program (NAIP) [Image collection]. Retrieved from: https://code.earthengine.google.com/

US Geological Survey Earth Resources Observation and Science Center. (2014). *Landsat TM Surface Reflectance* [Image collection]. Retrieved from https://dx.doi.org/10.5066/F78S4MZJ

US Geological Survey Earth Resources Observation and Science Center. (2014). *Provisional Landsat OLI Surface Reflectance*. [Image collection]. Retrieved from https://dx.doi.org/10.5066/F78S4MZJ

This material contains modified Copernicus Sentinel data (2015 - 2016), processed ESA.