Great Lakes Water Resources II

A Google Earth Engine Tool to Automate Wetland Mapping Using Optical and Radar Satellite Sensors in the Great Lakes Basin for Wetland Management and Monitoring

**Technical Report**

Final Draft – April 3rd, 2020

Vanessa Valenti (Project Lead)

Erica Carcelen

Katie Lange

Nicholas Russo

Bruce Chapman, NASA Jet Propulsion Laboratory, California Institute of Technology (Science Advisor)

Benjamin Holt, NASA Jet Propulsion Laboratory, California Institute of Technology (Science Advisor)

Previous Contributors:

Erica O’Connor

Melissa Ferriter

Alice Lin

Christopher Notto

# 1. Abstract

The Great Lakes Basin is one of the world’s largest freshwater ecosystems. The Basin harbors over 200,000 acres of wetlands. These wetlands provide a variety of environmental, ecological, and recreational functions to over 30 million people in the region. Some of these functions include improving water quality, mitigating flood impacts, providing wildlife habitat, and housing recreational activities. However, due to anthropogenic activities, habitat conversion and degradation threaten to disrupt or destroy remaining wetland ecosystems. Maps of wetland distribution based on ground surveys are costly and labor-intensive, prohibiting timely evaluations of wetland loss and gain. The Great Lakes Water Resources II team at the NASA Jet Propulsion Laboratory developed the Wetlands Extent Tool 2.0 (WET 2.0) in Google Earth Engine to automate mapping of wetland distribution in the Great Lakes Basin. The team partnered with the US Fish and Wildlife Service (USFWS), Environmental Protection Agency (EPA), Minnesota Department of Natural Resources (MDNR), the National Oceanic and Atmospheric Administration (NOAA), and Ducks Unlimited (DU). WET 2.0 incorporates Landsat 8 Operational Land Imager (OLI), Sentinel-1 C-band Synthetic Aperture Radar (C-SAR), and Sentinel-2 Multispectral Instrument (MSI) satellite data. WET 2.0 is trained to classify anywhere in the Great Lakes Basin. Utilizing a Random Forest classifier, WET 2.0 is capable of automatically mapping wetland extent in the entire Great Lakes Basin, achieving a mean overall accuracy of 80.12% when tested in Michigan. Findings and maps produced in WET 2.0 will enable our partners to identify areas of ecosystem degradation and wetland destruction in order to enact environmental practices and policy initiatives to maintain environmental and economic health in the area.

**Keywords**

Google Earth Engine, wetland, land-type classification, random forest, Synthetic Aperture Radar

# 2. Introduction

* 1. ***Background Information***

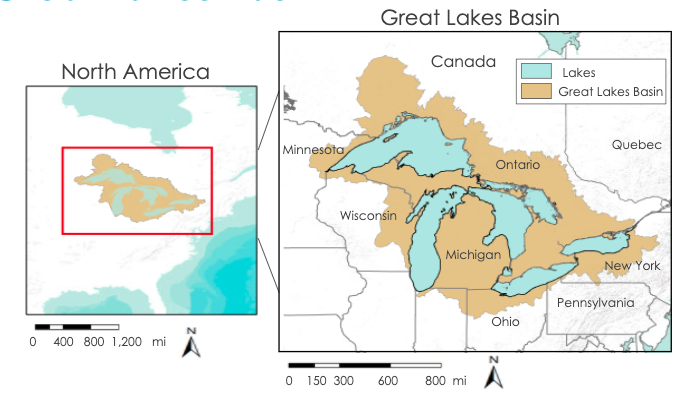
Wetlands provide crucial ecological and environmental functions, including flood and storm mitigation, coastal protection, wildlife habitat, hydrologic connectivity, sediment retention and stabilization, carbon sequestration, and water conservation (Mitsch et al., 2013; Mahdianpari et al., 2019). However, wetland habitats are changing and disappearing due to anthropogenic activities and natural processes. Agricultural and industrial development, water diversion, and changing precipitation patterns threaten wetlands worldwide (Tiner et al., 2015; Mahdianpari et al., 2019).

The subsequent need to monitor and restore wetlands has been recognized in recent decades. In 1986, the United States Congress identified wetlands as a nationally important resource and required updates of the nation’s wetland status and trends at 10-year intervals (Dahl, 2011). As a result, the US Fish and Wildlife Service (USFWS) now maintains a National Wetland Inventory (NWI) in order to generate wetland maps that inform stakeholders on wetland change, extent, and characteristics (US Fish and Wildlife Service, 2018). Historically, wetlands have been mapped through ground-based survey techniques and orthophotography. Though field monitoring is highly informative of wetland status, in situ data collection is expensive and time-consuming. In addition, many wetlands are in remote areas or logistically challenging to survey (Gallant, 2015). For these reasons, monitoring wetlands remotely via satellite data has become an attractive approach for long-term, comprehensive wetland delineation.

Several wetland characteristics pose remote sensing challenges; their highly dynamic nature, lack of unifying land cover features, and high potential for cloud cover can limit data availability and quality. The remote sensing community has addressed these concerns by incorporating complementary optical and radar datasets that improve delineation of spectrally similar wetland types (Mahdianpari et al., 2020). Optical data is sensitive to the chemical and molecular structure of vegetation while radar data, such as Synthetic Aperture Radar (SAR), is sensitive to the geometric and physical structure of vegetation (Mahdianpari et al., 2020). SAR data is especially useful to determine the flooding status of vegetation and remains unaffected by cloud cover or day/night conditions (Mahdianpari et al., 2020). Synthesizing optical and radar datasets allows for a comprehensive classification of wetland delineation independent of cloud cover and nightfall.

Traditionally, generating large-scale image composites and executing advanced classification algorithms required massive data storage capacity and high computational efficiency (Mahdianpari et al., 2019). Recently, the development of cloud-based computational frameworks such as Google Earth Engine (GEE), the availability of open-access Earth observation (EO) data, and advancements in machine learning techniques have made comprehensive, large-scale wetland delineation mapping attainable. GEE provides access to satellite data on a planetary scale and offers extensive computing power for image processing and analysis (Ahmad et al., 2019). Through these recent advances, successful studies have harnessed the power of GEE to process and analyze optical and SAR data for wetland delineation.

The Spring 2019 NASA DEVELOP Great Lakes Water Resources I project initially developed a Wetland Extent Tool (WET 1.0) in GEE for the state of Minnesota. They focused on mapping the growing season of 2017 (May through September). The team did so using a multi-source, multi-temporal, and object-based random forest classification approach with satellite datasets from Sentinel-1 C-Band Synthetic Aperture Radar (C-SAR) and Landsat 8 Operational Land Imager (OLI).



*Figure 1.* Study Area of the Great Lakes Basin in the Northeast Region of the United States and Canada

The Spring 2020 Great Lakes Water Resources II team focused on expanding WET 1.0 to encompass the entire Great Lakes Basin (Figure 1) and incorporate Sentinel-2 Multispectral Instrument (MSI) to increase wetland extent accuracy and resolution. The study period was determined to be the growing season of 2019 (May through September) as the region of interest was not frozen during this time. WET 2.0 will allow partners to easily access frequently updated, comprehensive, and accurate wetland extent maps to inform decisions regarding wetland conservation and restoration.

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

End-users of WET 2.0 include the USFWS National Wetlands Inventory (NWI), Environmental Protection Agency (EPA), Minnesota Department of Natural Resources (MN DNR), and Ducks Unlimited. Additional collaborative partners involved in the development of the tool include the University of Minnesota, Michigan Technological University, National Oceanic and Atmospheric Administration (NOAA) Office for Coastal Management, and Natural Resources Canada (NRCan). Partners will be able to use the final product to enhance existing research, assess the feasibility of future research, and inform end-user decisions concerning wetland management.

The tool provides an automated platform that uses Earth observing satellites to update wetland extent for the entire Great Lakes Basin. WET 2.0 allows for user-specified spatiotemporal analysis of wetlands in the Great Lakes Basin. This will enable partners to accurately monitor, evaluate, and analyze wetland extent and type in order to address environmental challenges in a timely and cost-effective manner.

# 3. Methodology

***3.1 Data Acquisition***

We accessed optical and radar satellite data through the GEE data catalog, including Landsat 8 OLI Level 2 Surface Reflectance (SR) Tier 1, Sentinel-2 MSI Level 1C Top of Atmosphere (TOA) Reflectance, and Sentinel-1 C-SAR Level-1 Ground Range Detected (GRD) products from 2019 (Table 1). For topographic data, we also used the GEE data catalog to access the Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM) from 2000. We decided to use the Sentinel-2 MSI Level 1C TOA products rather than Sentinel-2 MSI Level 2 SR data, which has been pre-processed by GEE primarily for general terrain correction, due to backlogs in the GEE data catalog. We automated an atmospheric correction algorithm for Level 1C data in GEE (described in 3.2) to extend Sentinel-2 data coverage to 2017 within WET 2.0.

Table 1

*Remote sensing data acquired in GEE used for WET 2.0 inputs*

|  |  |  |  |
| --- | --- | --- | --- |
| **Sensor** | **Processing Level** | **Data Provider** | **GEE ImageCollection ID** |
| Landsat 8 OLI | Level 2 SR Tier 1 | United States Geological Survey (USGS) | LANDSAT/LC08/C01/T1\_SR |
| Sentinel-2 MSI | Level 1C TOA Reflectance | European Space Agency (ESA Open Access Hub) | COPERNICUS/S2 |
| Sentinel-1 C-SAR | Level 1 GRD | European Space Agency  (ESA Open Access Hub) | COPERNICUS/S1\_GRD |
| SRTM | Level 2 Version 3 | NASA, USGS,  JPL - Caltech | USGS/SRTMGL1\_003 |

Additionally, we included ancillary datasets of land cover and land use in the US and Canada, USGS shapefiles of the study area, and field data points provided by our partners (Table 2). WET 2.0’s base shapefile for the study area came from the USGS’s Great Lakes Sub Basins shapefile, which divides the basin into 5 sub basins for reference. Field data was provided by our partners at the Michigan Technological University consisting of 681 field data points of wetlands in the Great Lakes Basin, collected 2017-2019, and our partners at Natural Resources Canada consisting of 125 polygons in Ontario.

Table 2

*Ancillary datasets incorporated into WET 2.0*

|  |  |  |  |
| --- | --- | --- | --- |
| **Data** | **Specifications** | **Source** | **GEE ImageCollection ID** |
| Cropland Data Layer | Raster | United States Department of Agriculture, National Agricultural Statistics Service (USDA NASS) | USDA/NASS/CDL |
| Canada Annual Crop Inventory | Raster | Agriculture and Agri-Food Canada (AAFC) | AAFC/ACI |
| Great Lakes Subbasins Shapefile, HUC 08 | Shapefile | USGS, Great Lakes Restoration Initiative | N/A |
| Wetland Identification Field Data Points | Data points | Michigan Technological University; Canada Natural Resources | N/A |
| North America Land Change Monitoring System (NALCMS) | Raster | Natural Resources Canada, Canada Centre for Remote Sensing, USGS | N/A |

***3.2 Data Processing***

We used the Google Earth Engine API to create the WET 2.0 tool and process all data in our study. The WET 2.0 tool is designed to create wetland classification maps for a user-defined time period and in a user-defined study area within the Great Lakes Basin. However, our analysis focused on the regions of interest to our partners during 2019. All data were clipped to the study area and filtered to represent the 2019 growing season. Imagery was projected and resampled to ensure the images align.

*3.2.1 Data Correction and Pre-processing*

SAR data in our study were filtered for Interferometric Wide mode, which is more suitable for wetland studies (Mahdianpari et al., 2020). We did not apply corrections to the radar data because the scenes available in GEE have already been processed in the European Space Agency Sentinel-1 toolbox, which includes terrain correction, applied orbit file, noisy border removal, radiometric calibration, and conversion from natural values to decibels (dB) (Google Developers, 2019). We converted the values from dB back to natural values (Equation 1) to ensure our raster calculations were accurate. The resulting VV and VH backscatter intensity image collections were then aggregated to create annual composites representing the mean values.

(Equation 1)

The optical imagery we used included Landsat 8 OLI and Sentinel-2 MSI. Optical imagery can be impeded by cloud cover and atmospheric haze, which need to be corrected and removed. No additional corrections were needed for the Landsat data because the Level-2 collections in GEE already have radiometric and geometric corrections applied. To remove the impact of clouds, we used bits 3 and 5 from the Quality Assessment (QA) band to mask out clouds and cloud shadows in Landsat 8 data.

For the Sentinel-2 MSI data, the improved dark object subtraction technique was employed for atmospheric correction of Level 1C images and adopted into GEE (Chavez, 1988). This technique removes atmospheric scattering, or “haze,” from top-of-atmosphere reflectance values to produce images representing surface reflectance. Top-of-atmosphere reflectance of each band was first converted to radiance. Each band in an image has an associated haze value, which is assumed to be the darkest value in the image. Due to the correlation among raw haze values for each band, the improved technique calculates a relative haze value for each band, starting with an initial haze value from the red band and standardizing based on band wavelength. We subtracted band-specific haze values from band-specific radiance and converted resulting radiance values to surface reflectance. In order to verify the reliability of the dark object subtraction for removing atmospheric scattering effects, we performed a series of Pearson’s correlation tests between mean composites of the three indices produced by the corrected Level 1C data and Level 2 SR data. Both data sets were filtered by date to represent the 2019 growing season and clipped to the shape of the Fond du Lac Reservation in northern Minnesota. Cloud masking information in the Sentinel-2 MSI data is stored in a quality assessment bitmask band named QA60. We used bits 10 and 11 in QA60 to mask out opaque and cirrus clouds, respectively.

When using remote sensing and satellite data to map wetland extent, in situ data that documents the geographical boundaries of wetlands is crucial information. Additionally, since WET 2.0 utilizes polygons to conduct analyses over the study period, it is important that our training and validation data are also polygons. Field data collected and provided by Michigan Technological University documented wetlands at least 40 x 50 m (0.2 hectare) in size and assessed wetland class, function, and health, represented as data points at the center of a wetland. The team took a conservative approach and buffered the field data points to create polygons representing an area of 0.1 hectares. The field data provided by NRCan was already in polygon format, so we only reclassified the data to fit our three-class system. The reclassification details for both field datasets are detailed in Appendix A. To supplement classes with fewer polygons, the team digitized Open Water and Upland polygons using the satellite base layer, slope calculated from elevation data, and the North American Land Change Monitoring System (NALCMS) land cover dataset as reference. The resulting 1,482 polygons (512 upland, 376 open water, and 594 wetland) were used for classifier training and data validation.

Several masks were employed to prevent class confusion and minimize processing time. In the SAR data, wetlands should have a double bounce scattering mechanism because the transmission wavelength can bounce off the water’s surface and standing vegetation perpendicular to the water’s surface. Urban land cover types also experience a double bounce scattering mechanism as transmissions bounce off the ground and then perpendicular buildings, which results in similar backscatter signatures in the SAR data. We masked out urban land cover using the NALCMS land cover dataset to avoid class confusion and improve accuracy. Agriculture can also present similar signatures to wetlands due to irrigation and growing practices, such as cranberry bogs. In order to differentiate wetlands from agriculture, we masked cropland data derived from the annual United States Department of Agriculture, National Agricultural Statistics Service (USDA NASS) and Agriculture and Agri-Food Canada (AAFC) crop inventory maps. The WET 2.0 tool accesses the crop inventory map from the year closest to the input date and includes all agriculture classes in the mask (Appendix A).

*3.2.2 Classification inputs*

Our classification model used a variety of inputs and indices designed to identify parameters characteristic of wetlands. These inputs include the Tasseled Cap Wetness Greenness Difference Index (TCWGD), Modified Normalized Difference Water Index (MNDWI), Normalized Difference Vegetation Index (NDVI), Dynamic Surface Water Extent (DSWE), VV backscatter, VH backscatter, and VV/VH ratio, which are further described in the following paragraphs. The annual composites for each input were then stacked into a multi-band image to undergo classification.

Our tool used Sentinel-2 MSI optical data to calculate TCWGD (Table 3; Shi & Xu, 2019), MNDWI (Equation 2), and NDVI (Equation 3), which are commonly utilized indices for wetland classification due to their ability to distinguish vegetation and water characteristics from other land cover types (Gao, 1996; Rouse Jr., Haas, Schell, Deering, & Harlan, 1974). The Red, Near Infrared (NIR), Green, Blue, Short Wave Infrared (SWIR), and Short Wave Infrared 2 (SWIR2) bands from the user-specified dates were used to calculate these indices. Additionally, the optical datasets were combined with topographic data from SRTM to calculate DSWE. Although topographic inputs have been shown to improve wetland classifications because of their relationship with water bodies, topography itself does not change frequently.

Table 3

*Tasseled cap transformation coefficients*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Index** | **Blue** | **Green** | **Red** | **NIR** | **SWIR** | **SWIR2** |
| Greenness | -0.3599 | -0.3533 | -0.4734 | 0.6633 | 0.0087 | -0.2856 |
| Wetness | 0.2578 | 0.2305 | 0.0883 | 0.1071 | -0.7611 | -0.5308 |

Meanwhile, wetland conditions and inundation change frequently, which can be captured in an index like DSWE that combines stationary topographic information with dynamic vegetation, soil, and wetness parameters (Jones, 2019). DSWE employs an algorithm that applies five decision rule-based diagnostic tests (Table A1). This algorithm determines if a pixel is fully covered by water or detects inundation in the presence of non-water land covers. The pixel is then assigned to one of six classes: not water, water (high confidence), water (moderate confidence), potential wetland, water or wetland (low confidence), or NA. (Table A2) (Ahmad, Hossain, Eldardiry, & Palveskey, 2019; Jones, 2019).

Radar data helped identify inundated wetlands, particularly when clouds are present, using the VV backscatter values, VH backscatter values, and VV/VH ratio. Open water should have low backscatter values due to smooth surface scattering and inundated wetlands should have high backscatter values due to double bounce scattering (Ahmad, Hossain, Eldardiry, & Palveskey, 2019). The VV/VH ratio is particularly useful for wetland classification because it is sensitive to soil moisture and corrects for terrain effects.

*3.2.3 Random Forest Classification*

The WET 2.0 tool approached wetland classification through the random forest (RF) method. The RF classification algorithm is a non-parametric machine learning technique that creates multiple decision trees based on training data sites and bands of an input image stack (Bourgeau-Chavez, Endres, Battaglia, Miller, Banda, Lauback, Higman, Chow-Fraser & Marcaccio, 2015; Mahdianpari et al, 2019). This method is frequently used in the literature for wetland mapping because it can handle a dataset that stems from multiple sources, has few observations, and is not normally distributed while remaining insensitive to overfitting and noise (Bourgeau- Chavez et al, 2015; Mahdianpari et al, 2019; Rodriguez-Galiano, Ghimire, Rogan, Chica-Olmo & Rigol-Sanchez, 2012). Additionally, the previous term found that the RF classifier produced a result with higher accuracy than the result created using a thresholding technique.

In order to ensure the WET 2.0 tool could classify anywhere in the Great Lakes Basin, regardless of whether there was field data present in the user-specified area, we incorporated two methods to the classification process. The previous term of this project filtered the field data to those within the user-specified area, but the classifier cannot run if there is no field data present. To address this issue, we filtered the field data based on the lake basin corresponding to the user-specified area. The WET 2.0 tool calculated the centroid of the user-specified area, identified the lake basin where the centroid was located, and selected all the field data within that basin. The second new approach we employed was creating two separate image stacks, one for training and one for classification, that both consisted of the composited inputs described in Section 3.2.2. Due to time constraints, we used the Level 2 product for the indices calculated with Sentinel-2, rather than the Level 1 product with our atmospheric correction applied, to create the training image stack for the entire 2019 year throughout the whole Great Lakes basin. The resulting training stack was imported into the WET 2.0 tool and the classification image stack was created within the WET 2.0 tool for the user-specified area and dates.

We employed 50% of the filtered field polygons and the values for each parameter in the training image stack to train the RF classifier. The trained RF classifier was then applied to the classification stack to create an initial classification identifying upland, open water, and wetland. RF uses the decision trees it generates to determine classification by selecting the class with the most votes among all the trees. We then applied the urban and agriculture masks, automatically reclassifying these areas as upland, to produce the final wetland classification maps.

***3.3 Data Analysis***

Using field data provided by our partners at Michigan Technological University, we utilized 50% of the 681 field polygons collected in the Great Lakes Basin in 2019 for validation against WET 2.0’s outputs. WET 2.0 output classification map layers were randomly sampled to generate “predicted” values (wetland classification or presence) for comparison against the field “actual” values. We then generated a confusion matrix and conducted an accuracy assessment in GEE. We also computed the Kappa coefficient, consumer’s accuracy, and producer’s accuracy in GEE for additional statistical analysis. The overall accuracy assessment indicates what percentage of field data points were correctly classified by WET 2.0, while the Kappa coefficient evaluates the classification’s performance compared to a random classification of pixels. Consumer’s accuracy measures how often the class on the map will actually be present on the ground (reliability), while producer’s accuracy measures how often real features on the ground are correctly shown on the classified map.

# 4. Results & Discussion

***4.1 Analysis of Results***

*4.1.1. Dark object subtraction evaluation*

The team developed and applied an atmospheric correction of Sentinel-2 Level 1C images in GEE using dark object subtraction. All three indices produced from corrected Level 1C images -- NDVI, MNDWI, and TCWGD -- correlated significantly with those from Level 2 SR images, indicating that our dark object subtraction method resulted in sufficiently similar outputs suitable for analysis in our tool (r >0.93 for all comparisons; p<0.0001 for all comparisons; Table 4). We were thus confident using an automated atmospheric correction of dark object subtraction for Sentinel-2 images in WET 2.0.

Table 4

*Pearson’s correlations between Sentinel-2 Level 1C and Level 2 SR images for three indices*

|  |  |  |
| --- | --- | --- |
| **Index** | **Comparison** | **Pearson’s correlation coefficient (*r*)** |
| TCWGD | Level 1C corrected – Level 2 SR | 0.965 |
| MNDWI | Level 1C corrected – Level 2SR | 0.933 |
| NDVI | Level 1C corrected – Level 2SR | 0.946 |

*4.1.2. Classification Evaluation*

Our team produced five classification maps, one for each lake basin in the Great Lakes region (Appendix B). The classification maps for May - June 2019 achieved a high mean overall accuracy of 93% when compared to the reserved validation dataset. Additionally, the mean kappa value for our classification maps was 87, indicating our model could classify wetlands better than a random model (Table 5). The classification for the Lake Superior basin had the highest overall accuracy and the Lake Ontario basin had the lowest overall accuracy. Although Lake Superior had the highest overall accuracy, less than 10% of the wetland areas identified by the field data were classified correctly and only 32% of the area classified as wetland were actually representative of wetlands (Table 5). The high overall accuracy in Lake Superior is likely relying on correct classification of upland and open water areas, as indicated by values greater than 98% for the consumer’s and producer’s accuracies in both classes. Additionally, while the Lake Ontario basin classification had the lowest overall accuracy, most (> 69%) of the areas classified as wetland were actual wetlands in the field data. However, the Lake Ontario classification missed many wetland areas since it correctly classified only 3% of the wetlands in the field data (Table 5). These discrepancies highlight the potential for overall accuracy values to misrepresent a classifier’s output and the need to examine multiple accuracy measures.

The Lake Michigan basin classification, which had the second-highest overall accuracy, had the second-highest consumer’s accuracy (44%) and the lowest producer’s accuracy (15%) when comparing among the basin classifications. This indicates the Lake Michigan basin classification correctly classified 44% of the actual wetland areas by over-classifying wetland pixels. The Lake Huron basin classification under classified wetlands as most (56%) of the area classified as wetland were actual wetlands and less than a third of the actual wetlands in the validation dataset were identified (Table 5). Meanwhile, the high overall accuracy of the Lake Erie basin classification is likely the most reliable since most (58%) of the actual wetlands in the validation dataset were correctly classified and most (51%) of the area classified as wetlands were actually wetlands. This would indicate our classification model performed best in the Lake Erie basin.

While the producer and consumer accuracies for each class evaluate how well the classification model identifies each class, confusion matrices (Appendix B) help to understand class confusion. In all five basins, the actual wetland areas were most often misclassified as upland, which was most apparent in the Lake Ontario and Lake Superior basins. Additionally, areas misclassified as wetlands were most often actually uplands in the validation dataset in the Lake Huron, Lake Erie, and Lake Michigan basins. Most of the areas classified as wetlands were misclassified in the Lake Superior and Lake Michigan basins, while more than 50% of the areas classified as wetlands in the Lake Huron, Lake Ontario, and Lake Erie basins were correctly classified.

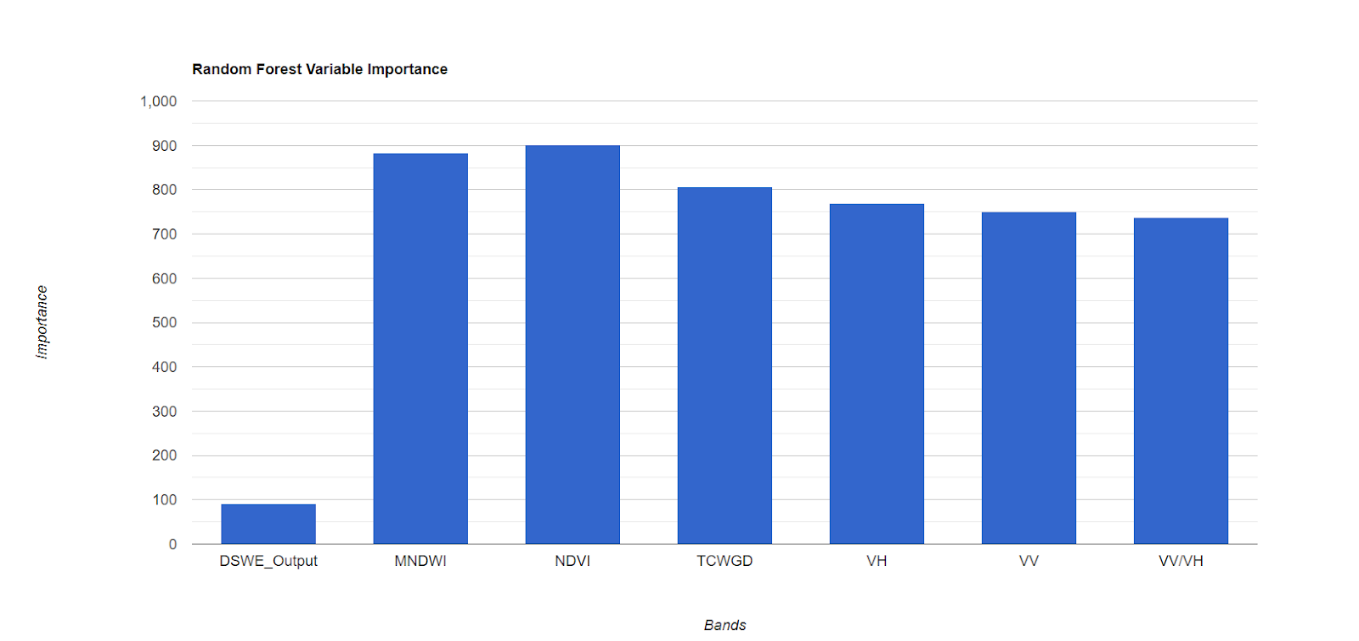
Table 5

*Accuracy Measures (%) for All Basin Classifications*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Lake Basin** | **Overall Accuracy** | **Kappa** | **Consumer’s Accuracy** | | | **Producer’s Accuracy** | | |
| **Upland** | **Open Water** | **Wetland** | **Upland** | **Open Water** | **Wetland** |
| Superior | 99.81% | 98.97% | 99.92% | 99.90% | 9.66% | 98.43% | 99.99% | 32.86% |
| Huron | 93.33% | 87.89% | 97.74% | 98.78% | 29.05% | 89.64% | 99.99% | 56.56% |
| Ontario | 77.43% | 63.61% | 99.86% | 98.87% | 2.94% | 58.30% | 99.55% | 69.56% |
| Erie | 94.75% | 88.27% | 96.08% | 99.04% | 58.72% | 96.76% | 99.93% | 51.95% |
| Michigan | 97.59% | 94.32% | 96.00% | 98.81% | 44.05% | 98.22% | 99.95% | 15.37% |
| **All Basin Mean** | **92.58%** | **86.61%** | **97.92%** | **99.08%** | **28.89%** | **88.27%** | **99.88%** | **45.26%** |

*4.1.3. Variable Importance*

Each basin classification has a corresponding variable importance graph from which we were able to identify general trends. As an example, the variable importance graph for the Lake Erie basin classification is shown in Figure 4, and all of the graphs for the individual basins are included in Appendix C. The most important variables were NDVI, MNDWI, and TCWGD, which consistently ranked in the top three for all five basin classifications. The VH band ranked as the fourth most important variable in the Lake Ontario, Lake Erie, and Lake Michigan basin classifications. The other two radar inputs, VV and the VV/VH ratio, ranked in the 4th to 6th positions in all five basins classifications. DSWE, which was the parameter added this term, was ranked as the least important variable in all five basin classifications. These results indicate our optical indices were the most useful for wetland classification and DSWE was the least useful. DSWE was the only categorical variable, which could explain its low ranking in terms of variable importance.



*Figure 4.* Variable importance for the Lake Erie (May-June 2019) basin classification

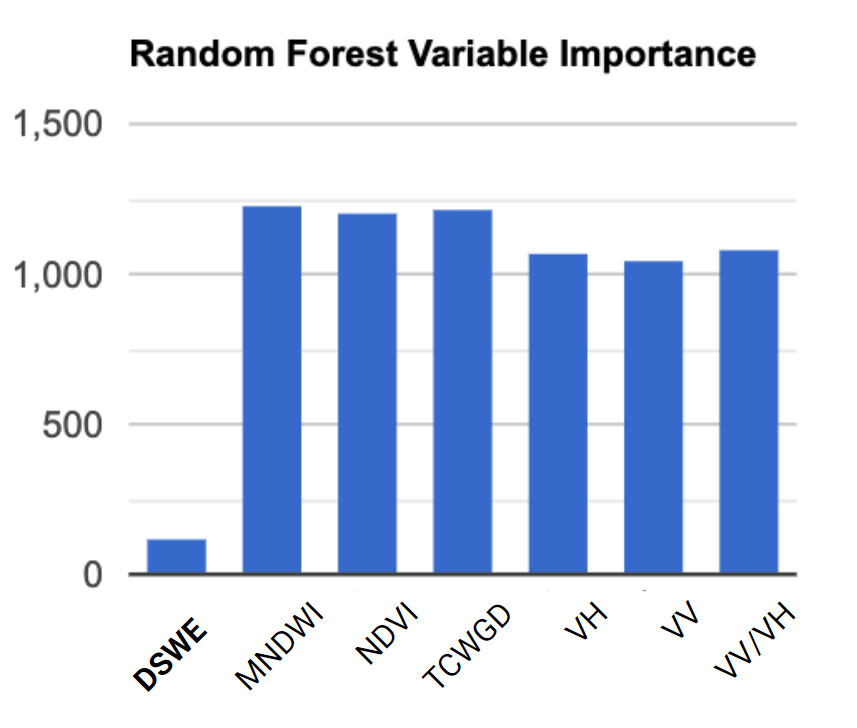
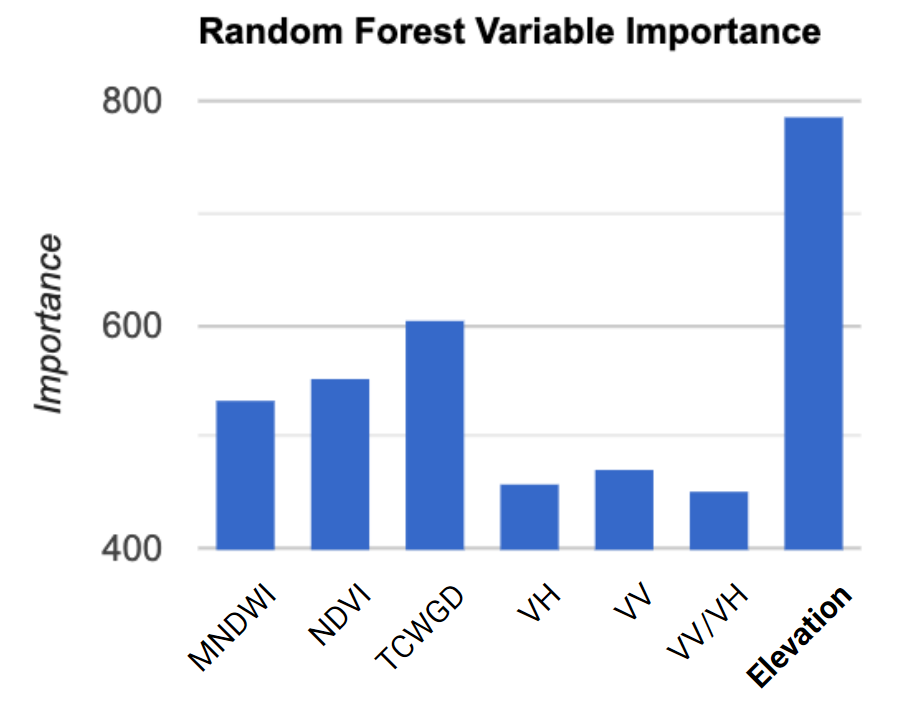
*4.1.4. Elevation and DSWE*

DSWE was the parameter in our model that incorporated topographic information by utilizing slope and hillshade calculated from elevation within its five decision rule diagnostic tests. In order to explore how DSWE was affecting our classification, we ran two classifications in northern Michigan during the 2019 growing season (Appendix D). One classification model had the seven parameters detailed in Section 3.2.2, while the other classification model used these seven parameters with elevation rather than DSWE. The model using elevation appears more accurate with more wetlands classified and higher overall, producer’s, and consumer’s accuracy values (Table 6). Elevation was the most important variable when used instead of DSWE and reduces the importance of the six other parameters (Figure 5). This could be due to the test sites’ location along the coast, where elevation changes drastically. However, this could also indicate valuable topographic information was lost when incorporated through DSWE as an input for the RF classifier.

Table 6

*Accuracy Measures (%) for the Michigan Test Sites*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Run** | **Overall Accuracy** | **Consumer’s Accuracy** | | | **Producer’s Accuracy** | | |
| **Upland** | **Open Water** | **Wetland** | **Upland** | **Open Water** | **Wetland** |
| With Elevation | 99.62% | 99.66% | 100% | 76.67% | 99.97% | 99.28% | 30.26% |
| With DSWE | 99.56% | 99.60% | 100% | 70% | 99.95% | 99.64% | 25.30% |



*Figure 5.* Variable importance for each lake basin classification model.

***4.2 WET 2.0 Tool***

*4.2.1 Tool Development*

The Wetland Extent Tool (WET) was first developed by the previous Spring 2019 Great Lakes Water Resources team to identify wetland extent in Minnesota during the 2017 growing season. This term, the tool was expanded to encompass the entire Great Lakes Basin for the 2019 growing season. WET 2.0 improved accuracy and validation of wetland classification with the addition of high-resolution Sentinel-2 imagery. Sentinel-2 data was used to calculate NDVI, MNDWI, and TCWGD in place of Landsat 8 in order to increase accuracy in wetland classification. Additionally, Landsat 8 was utilized to calculate DSWE in the tool to further delineate wetlands. Lastly, Sentinel-1 remained the input for VV, VH, and VV/VH ratio. The tool automatically composites these calculated indices to output a single wetland classification map.

This term, the team developed a user interface for WET 2.0 in GEE. The tool’s user-friendly interface allows customizable wetland analysis for users with wide-ranging remote sensing experience. The user inputs a date range, selects an area of interest or draws a customizable polygon, and selects the desired parameter layers to be added to the map. WET 2.0 then outputs a filtered, masked, and corrected image collection for the selected parameter layers that are customized to the user-specified temporal and spatial inputs. These layers can then be exported to the user’s personal Google Drive for further analysis. Additionally, the interface has an option to generate a time series chart. When the user selects an analysis type and clicks a point on the map, a time series chart is generated to display the point values over the selected time frame.

*4.2.2 Tool Applications*

WET 2.0 allows end users to identify and monitor wetlands in the Great Lakes Basin with on-demand and accurate wetland maps. The tools user-friendly interface allows for spatially and temporally customizable wetland analysis. Partners will use these easily accessible, up to date, and accurate wetland maps to inform research and decision making in the Great Lakes Basin.

***4.3 Sources of Error***

It is important to note errors and uncertainties in WET 2.0 methodology in order to improve upon them in the future. One probable source of error in the tool is the image composition process. To classify wetlands in the tool, the input datasets are combined as a single final composite that the tool can classify. Averaging image composites results in loss of outlier information that may be useful in identifying wetlands. In addition, the training of WET 2.0 relies on field data. The field data used to train the tool consisted of many wetland polygons and few upland polygons. This has resulted in an over-representation of the wetland class in our training data and as a result, WET 2.0 may overestimate wetland extent in the Basin. Finally, in an effort to reduce pixel noise, the team planned to perform object-based image classification. However, due to the limited processing capacity of GEE, this was not possible for this project.

***4.2 Future Work***

In the future, very high-resolution optical data sets could be utilized for further classification and validation. High-resolution optical data could allow the delineation of more wetland classes and, if fine enough resolution, could be a proxy for field data validation. L-band SAR could be better utilized to penetrate farther into forested wetland to further delineate wetland types in the classification. This project could also benefit from an expanded study period to include past dates for improved wetland change detection and analysis, since this is particularly useful for partners.

# 5. Conclusions

The Great Lakes Water Resources II team successfully created a Google Earth Engine tool that is capable of automated wetland classification and mapping in the entire Great Lakes Basin. The team expanded on the original WET project and code and improved accuracy of the tool by utilizing 10-m resolution optical and radar data as opposed to 30-m data. WET 2.0 trained and validated against field data collected and provided by partners at Michigan Technological University and NRCan. The tool is capable of monitoring wetland extent in the entire Great Lakes Basin beginning in 2017 and updating as satellite data is processed in Google Earth Engine. WET 2.0 allows end-users to enter a time range and area of analysis and outputs map representations of wetlands according to the user’s specifications.

Project results will help inform decision makers and wetland conservation managers in the Great Lakes Basin. The tool will serve as a source of information to extent maps of wetlands and can be utilized to assess vegetation and water presence in the Great Lakes Basin by employing the multiple map layers and output options available to users in the tool. The team conducted an in-depth statistical validation and evaluation of the Random Forest classifier performance in the tool and concluded that it performs best in the Lake Erie Basin; more than 50% of classified wetlands in the Lake Erie, Michigan, and Huron Basins were correctly classified as indicated by the consumer’s accuracy in our field data validation. In our evaluation, we found that the field validation data or “actual” wetland values were most often misclassified as upland. The most important inputs to the tool’s classification were the optical indices, NDVI, MNDWI, and TCWGD, derived from Sentinel-2 10-m data, while DSWE was the least important variable in all results. The low importance of DSWE could indicate topographic information is lost, and may be utilized more effectively in methods that differ from our approach.

This term has greatly built the capacity of the WET tool for wetland classification so that users can refine the model to produce accurate classification maps for their area of interest. In addition to contributing to tracking and monitoring capabilities of wetland extent in the Great Lakes Basin, these results and analysis can aid in the understanding and use of optical and radar remote sensing data for the mapping of wetlands.

# 6. Acknowledgments

The Great Lakes Water Resources II team would like to thank our partners from the USFWS NWI, MN DNR, EPA, Ducks Unlimited, University of Minnesota, NOAA, Michigan Technological University, and NRCan for their feedback and involvement in the development of WET 2.0. Thank you to our science advisor, Dr. Bruce Chapman of NASA Jet Propulsion Laboratory for his guidance and expertise in this project. Thanks also to Benjamin Holt of NASA Jet Propulsion Laboratory and our NASA DEVELOP Fellow Cecil Byles for their help in project development, coordination, and general support throughout this project.

This material contains modified Copernicus Sentinel data (2016-2019), processed by ESA.

Any opinions, findings, and conclusions or recommendations expressed in this mupaterial 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.

# 7. Glossary

**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-based computing platform for planetary-scale environmental data analysis that is capable of performing highly-interactive algorithm development

**Ecosystem services/functions -** Ecological processes and use of ecosystems to benefit humans

**Random Forest (RF) classifier -** Ensemble learning method for classification that consists of a large number of individual decision trees

**SAR -** Synthetic Aperture Radar

**MSI -** Multispectral Instrument

**OLI -** Operation Land Imager

**GRD** - Ground Range Detected

**TOA** - Top of Atmosphere reflectance

**SR** - Surface Reflectance

**DEM** - Digital Elevation Model

**EPA -** Environmental Protection Agency

**US FWS -** United States Fish and Wildlife Service

**NOAA -** National Oceanic and Atmospheric Association

**NWI -** National Wetlands Inventory

**MN DNR** - Minnesota Department of Natural Resources

**NRCan** - Natural Resources Canada

**USGS** - United States Geological Survey

**ESA** - European Space Agency

**USDA** - United States Department of Agriculture

**NASS** - National Agricultural Statistics Service

**AAFC** - Agriculture and Agri-Food Canada

**QA** - Quality Assessment

**TCWGD** - Tasseled Cap Wetness Greenness Difference Index

**MNDWI** - Modified Normalized Difference Water Index

**NDVI** - Normalized Vegetation Index

**DSWE** - Dynamic Surface Water Extent

# SNIC - Simple Non-Iterative Clustering

# 8. References

Ahmad, S., Hossain, F., Eldardiry H., & Pavelsky. (2019). A fusion approach for water area classification using visible, near infrared and synthetic aperture radar for south Asian conditions. *IEEE Transactions on Geoscience and Remote Sensing*, 1-10. https://doi.org/10.1109/TGRS.2019.2950705

Bourgeau-Chavez, L., Endres, S., Battaglia, M., Miller, M., Banda, E., Laubach, Z., Higman, P., Chow-Fraser, P., & Marcaccio J. (2015). Development of a bi-national Great Lakes coastal wetland and land use map using three-season PALSAR and Landsat imagery. *Remote Sensing, 7*. https://doi.org/10.3390/rs70708655

Dahl, T. E. (2011). Status and trends of wetlands in the conterminous United States 2004 to 2009. Washington DC: US Department of the Interior, Fish and Wildlife Service. Retrieved from https://www.fws.gov/wetlands/documents/Status-and-Trends-of-Wetlands-in-the-Conterminous-United-States-2004-to-2009.pdf

Gallant, A. (2015). The challenges of remote monitoring wetlands. *Remote Sensing, 7*(8). 10938-10950. https://doi.org/10.3390/rs70810938

Gao, B. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment, 58*(3). https://doi.org/10.1016/S0034-4257(96)00067-3

Jones, J. (2019). Improved automated detection of subpixel-scale inundation - revised dynamic surface water extent (dswe) partial surface water tests. *Remote Sensing 11*, (4). https://doi.org/10.3390/rs11040374

Mahdianpari, M.., Salehi, B., Mohammadimanesh, F.., Brisco B., Homayouni, S., Gill, E., DeLancey, E., & Bourgeau-Chaves, L. (2020). Big data for a big country: The first generation of Canadian wetland inventory map at a spatial resolution of 10-m using Sentinel-1 and Sentinel-2 data on the Google Earth Engine cloud computing platform. *Canadian Journal of Remote Sensing, 46*(1), 15-33*.* https://doi.org/10.1080/07038992.2019.1711366

Mahdianpari, M., Salehi, B., Mohammadimanesh, F., Homayouni, S., & Gill, E. (2019). The first wetland inventory map of Newfoundland at a spatial resolution of 10m using Sentinel-1 and Sentinel-2 data on the Google Earth Engine Cloud Computing Platform. *Remote Sensing, 11* (1) https://doi.org/10.3390/rs11010043

Mitsch, W. J., Bernal, B., Nahlik, A. M., Mander, U., Zhang, L., & Anderson, C. J. (2013). Wetlands, carbon, and climate change. *Landscape Ecology, 28*, 583–597. doi: https://doi.org/10.1007/s10980-012-9758-8

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing, 67*. https://doi.org/10.1016/j.isprsjprs.2011.11.002

Rouse Jr., J. W., Haas, R. H., Deering, D. W., Schell, J. A., & Harlan, J. C. (1974). Monitoring the vernal advancement and retrogradation (greenwave effect) of natural vegetation. College Station, TX: Texas A&M University Remote Sensing Center. Retrieved from https://ntrs.nasa.gov/search.jsp?R=19750020419&hterms=Monitoring+Vernal+Advancement+Retrogradation+Greenwave+Effect+Natural+Vegetation&qs=N%3D0%26Ntk%3DAll%26Ntt%3DM onitoring%2520The%2520Vernal%2520Advancement%2520and%2520Retrogradation%2520%28G reenwave%2520Effect%29%2520of%2520Natural%2520Vegetation%26Ntx%3Dmode%2520match allpartial

Shi, T., & Xu, H. (2019). Derivation of tasseled cap transformation coefficients for Sentinel-2 MSI at-sensor reflectance data*. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12.* 4038-4048.

Tiner, R. W., Lang, M. W., Klemas, V. V., & Johnston, C. A. (2015). Remote Sensing of Wetlands: Applications and Advances. Natural Resources Management Faculty Books, 2. Retrieved from https://openprairie.sdstate.edu/nrm\_book/2/

US Fish & Wildlife Service. (2018). NWI Program Overview. Retrieved from https://www.fws.gov/wetlands/nwi/Overview.html

# 9. Appendices

Appendix A

Table A1

*DSWE Diagnostic Tests*

|  |  |  |
| --- | --- | --- |
| **Test ID** | **Threshold** | **Pixel Value** |
| 1 | MNDWI > 0.124 | 1 |
| 2 | (Green + Red) > (NIR + SWIR1) | 10 |
| 3 | AWEsh > 0.0 | 100 |
| 4 | MNDWI > -0.44 & SWIR1 < 900 & NIR < 1500 & NDVI < 0.7 | 1000 |
| 5 | MNDWI > -0.5 & SWIR1 < 3000 & SWIR2 < 1000 & NIR < 2500 & Blue < 1000 | 10000 |

Table A2

*DSWE Class Rules*

|  |  |  |  |
| --- | --- | --- | --- |
| **Class Value** | **Class Interpretation** | **Rule** | **Pixel Values** |
| 0 | Not Water | None true or one of Tests 1-4 are true | 0, 1, 10, 100, 1000 |
| 1 | Water  (high confidence) | Any four to five tests are true | 1111, 10111, 11011, 11101, 11110, 11111 |
| 2 | Water  (moderate confidence) | Any three tests are true | 111, 1011, 1101, 1110, 10011, 10101, 10110, 11001, 11010, 11100 |
| 3 | Potential wetland | Tests 4 and 5 are true | 11000 |
| 4 | Water or Wetland  (low confidence) | Any two tests are true or Test 5 is true | 11, 101, 110, 1001, 1010, 1100, 10000, 10001, 10010, 10100 |
| 255 | NA | Fill | NA |

Table A3

*Field Data Reclassification*

|  |  |  |
| --- | --- | --- |
| **Original Class** | **Reclass Value** | **Reclass Description** |
| Shrubby | 0 | Upland |
| Forest | 0 | Upland |
| Open Water | 1 | Open Water |
| Water | 1 | Open Water |
| Emergent | 2 | Wetland |
| Wet Meadow | 2 | Wetland |
| Bog | 2 | Wetland |
| Fen | 2 | Wetland |
| Wetland Shrub | 2 | Wetland |
| Marsh | 2 | Wetland |
| Forested Wetland | 2 | Wetland |
| Swamp | 2 | Wetland |

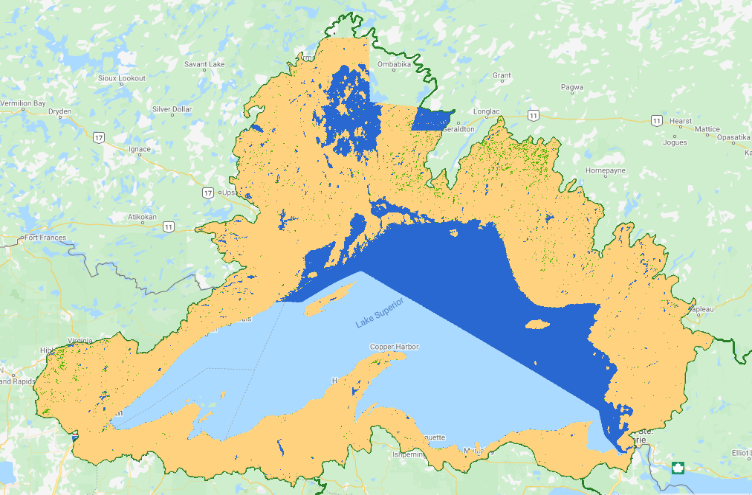
Table A4

*Urban and Agriculture Mask Values*

|  |  |
| --- | --- |
| **Mask** | **Classes Set to 1 (Unmasked)** |
| Urban – NALCMS | 17 |
| Agriculture – USDA Cropland | 63, 64, 65, 81, 82, 83, 87, 88, 92, 111, 112, 121, 122, 123, 124, 131, 141, 142, 143, 152, 176, 190, 195 |
| Agriculture – AAFC Cropland | 10, 20, 30, 34, 35, 50, 80, 110, 122, 130, 200, 210, 220, 230 |

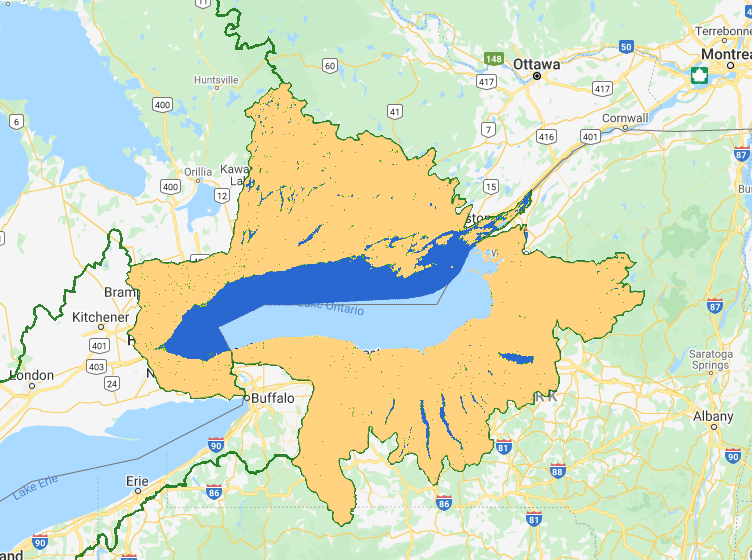
Appendix B

*Figure B1:* Wetland classification maps during May – June 2019 output by Wet 2.0 for each lake basin in the Great Lakes region.

a) Lake Superior b) Lake Huron



c) Lake Michigan d) Lake Ontario



e) Lake Erie

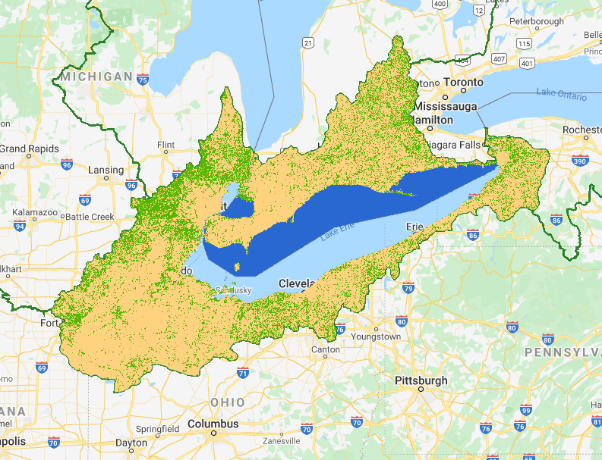




Table B1

*Confusion Matrices as Number of Pixels for Each Basin Classification*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lake Superior** | | **Actual** | | |
| **Upland** | **Open Water** | **Wetland** |
| **Predicted** | **Upland** | 23,628 | 179 | 197 |
| **Open Water** | 3 | 204,893 | 18 |
| **Wetland** | 16 | 31 | 23 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lake Huron** | | **Actual** | | |
| **Upland** | **Open Water** | **Wetland** |
| **Predicted** | **Upland** | 38,098 | 4 | 4,398 |
| **Open Water** | 0 | 41,176 | 5 |
| **Wetland** | 880 | 505 | 1,803 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lake Ontario** | | **Actual** | | |
| **Upland** | **Open Water** | **Wetland** |
| **Predicted** | **Upland** | 15,971 | 142 | 11,282 |
| **Open Water** | 0 | 23,770 | 107 |
| **Wetland** | 22 | 129 | 345 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lake Erie** | | **Actual** | | |
| **Upland** | **Open Water** | **Wetland** |
| **Predicted** | **Upland** | 16,797 | 3 | 559 |
| **Open Water** | 0 | 5,973 | 4 |
| **Wetland** | 686 | 55 | 801 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lake Michigan** | | **Actual** | | |
| **Upland** | **Open Water** | **Wetland** |
| **Predicted** | **Upland** | 25,914 | 90 | 379 |
| **Open Water** | 0 | 66,888 | 35 |
| **Wetland** | 1,079 | 716 | 326 |

Table B2

*Confusion Matrices as Consumer’s Accuracy (%) for Each Basin Classification*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lake Superior** | | **Actual** | | |
| **Upland** | **Open Water** | **Wetland** |
| **Predicted** | **Upland** | 99.92% | 0.09% | 82.77% |
| **Open Water** | 0.01% | 99.90% | 7.56% |
| **Wetland** | 0.07% | 0.02% | 9.66% |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lake Huron** | | **Actual** | | |
| **Upland** | **Open Water** | **Wetland** |
| **Predicted** | **Upland** | 97.74% | 0.01% | 70.87% |
| **Open Water** | 0.00% | 98.78% | 0.08% |
| **Wetland** | 2.26% | 1.21% | 29.05% |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lake Ontario** | | **Actual** | | |
| **Upland** | **Open Water** | **Wetland** |
| **Predicted** | **Upland** | 99.86% | 0.59% | 96.15% |
| **Open Water** | 0.00% | 98.87% | 0.91% |
| **Wetland** | 0.14% | 0.54% | 2.94% |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lake Erie** | | **Actual** | | |
| **Upland** | **Open Water** | **Wetland** |
| **Predicted** | **Upland** | 96.08% | 0.05% | 40.98% |
| **Open Water** | 0.00% | 99.04% | 0.29% |
| **Wetland** | 3.92% | 0.91% | 58.72% |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lake Michigan** | | **Actual** | | |
| **Upland** | **Open Water** | **Wetland** |
| **Predicted** | **Upland** | 96.00% | 0.13% | 51.22% |
| **Open Water** | 0.00% | 98.81% | 4.73% |
| **Wetland** | 4.00% | 1.06% | 44.05% |

Table B3

*Confusion Matrices as Producer’s Accuracy (%) for Each Basin Classification*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lake Superior** | | **Actual** | | |
| **Upland** | **Open Water** | **Wetland** |
| **Predicted** | **Upland** | 98.43% | 0.75% | 0.82% |
| **Open Water** | 0.00% | 99.99% | 0.01% |
| **Wetland** | 22.86% | 44.29% | 32.86% |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lake Huron** | | **Actual** | | |
| **Upland** | **Open Water** | **Wetland** |
| **Predicted** | **Upland** | 89.64% | 0.01% | 10.35% |
| **Open Water** | 0.00% | 99.99% | 0.01% |
| **Wetland** | 27.60% | 15.84% | 56.56% |

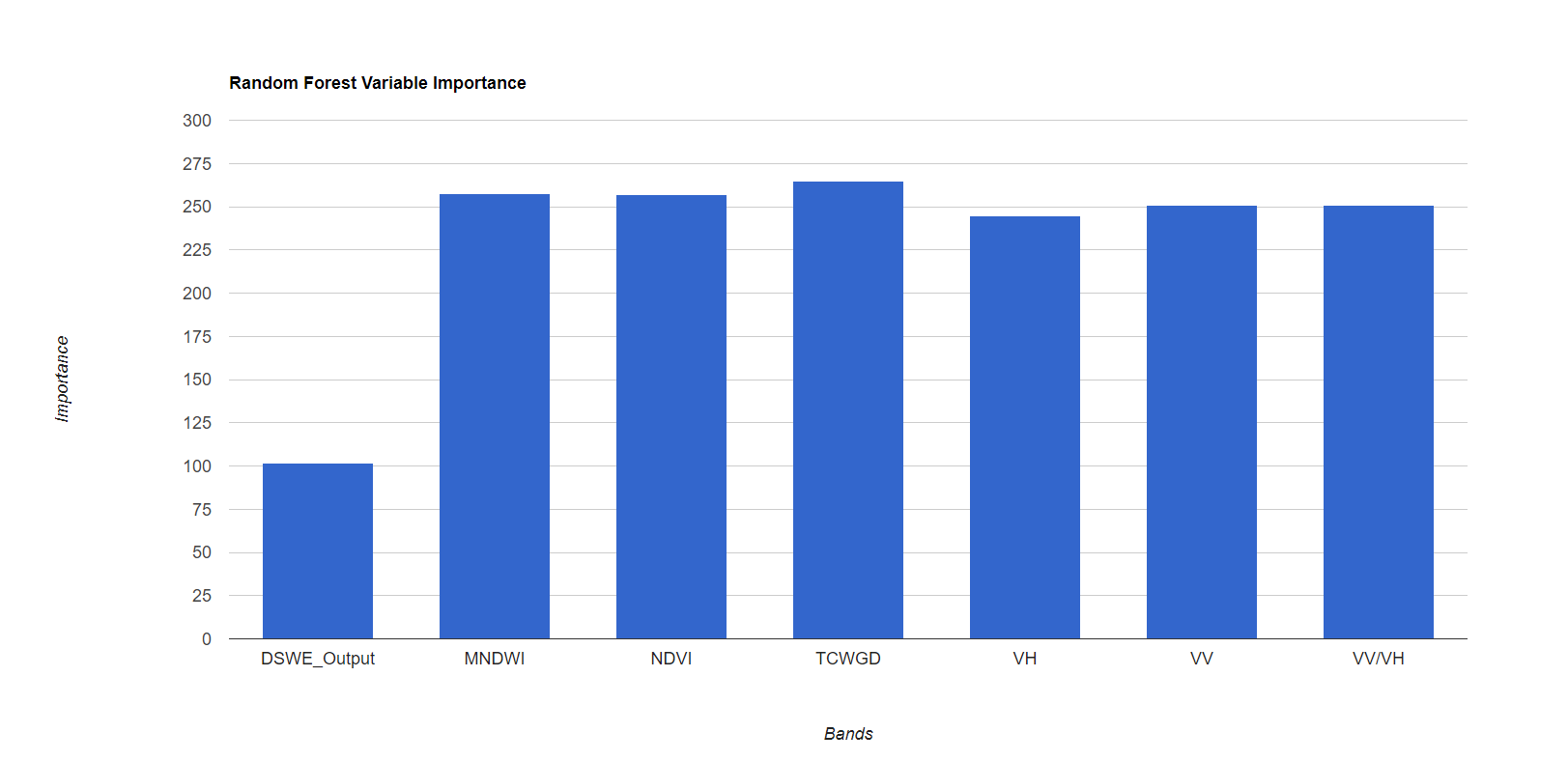
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lake Ontario** | | **Actual** | | |
| **Upland** | **Open Water** | **Wetland** |
| **Predicted** | **Upland** | 58.30% | 0.52% | 41.18% |
| **Open Water** | 0.00% | 99.55% | 0.45% |
| **Wetland** | 4.44% | 26.01% | 69.56% |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lake Erie** | | **Actual** | | |
| **Upland** | **Open Water** | **Wetland** |
| **Predicted** | **Upland** | 96.76% | 0.02% | 3.22% |
| **Open Water** | 0.00% | 99.93% | 0.07% |
| **Wetland** | 44.49% | 3.57% | 51.95% |

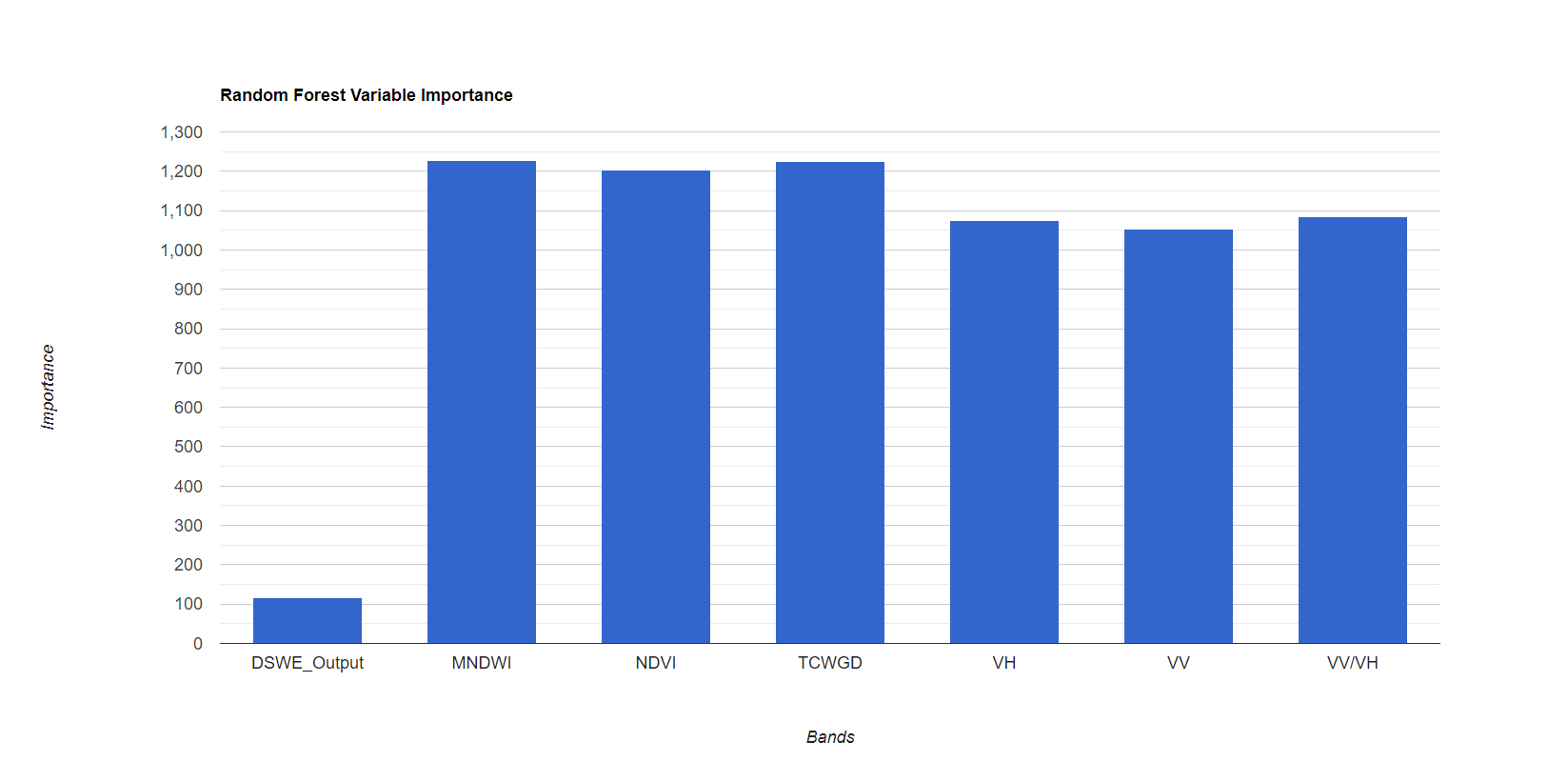
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lake Michigan** | | **Actual** | | |
| **Upland** | **Open Water** | **Wetland** |
| **Predicted** | **Upland** | 98.22% | 0.34% | 1.44% |
| **Open Water** | 0.00% | 99.95% | 0.05% |
| **Wetland** | 50.87% | 33.76% | 15.37% |

Appendix C

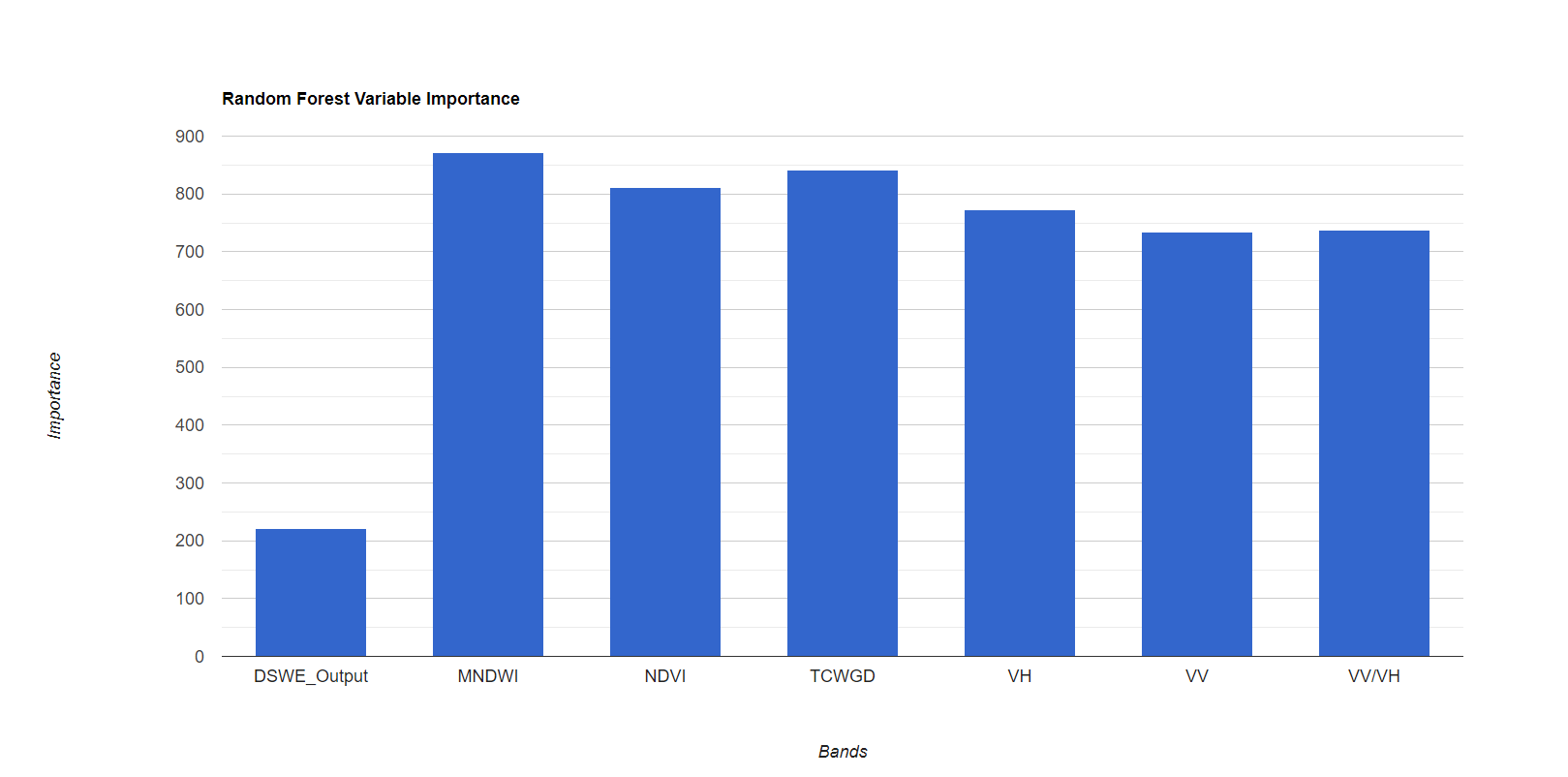
*Figure C1:* Variable Importance Graphs for the Lake Superior Basin Classification



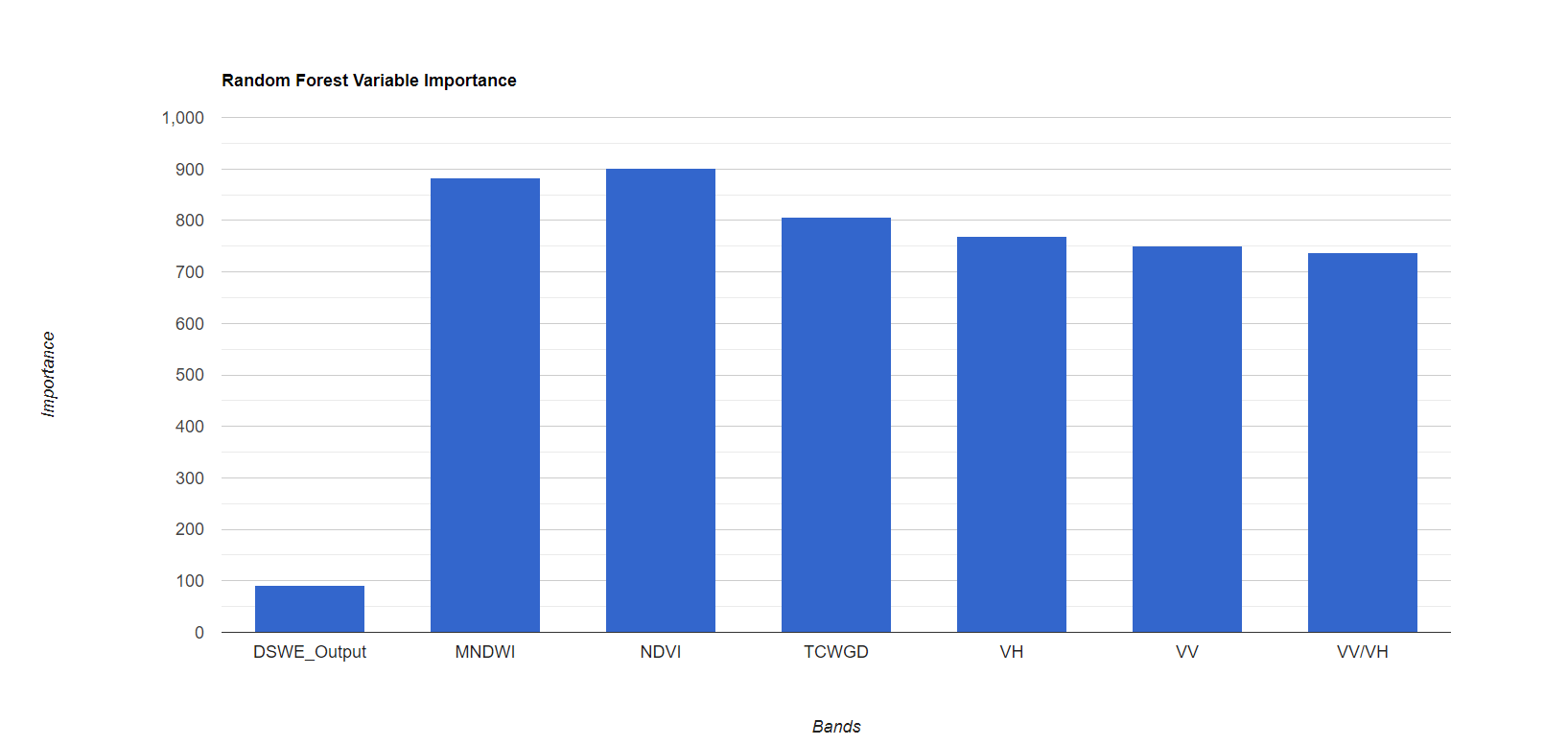
*Figure C2:* Variable Importance Graphs for the Lake Huron Basin Classification



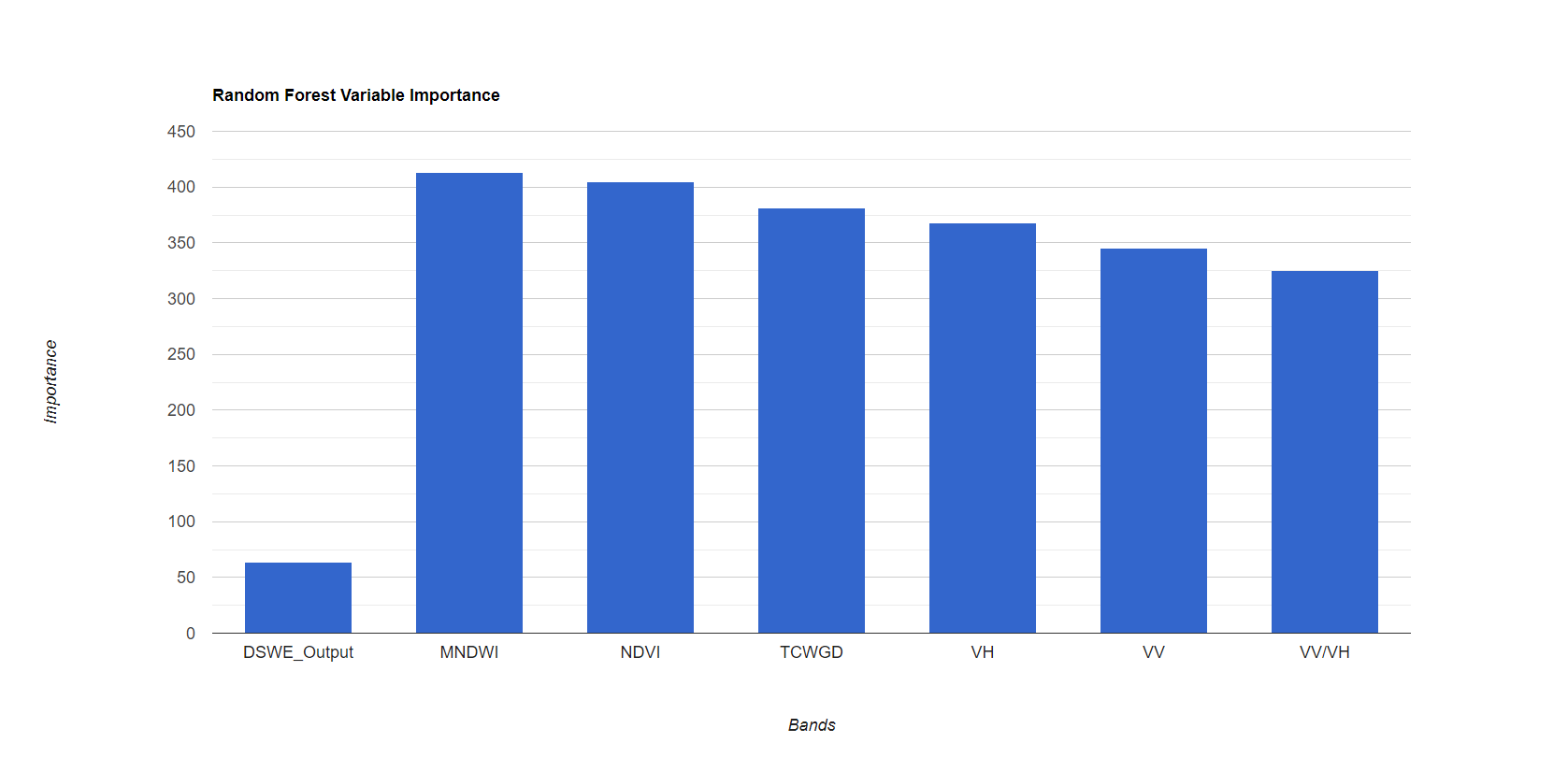
*Figure C3:* Variable Importance Graphs for the Lake Ontario Basin Classification



*Figure C4:* Variable Importance Graphs for the Lake Erie Basin Classification

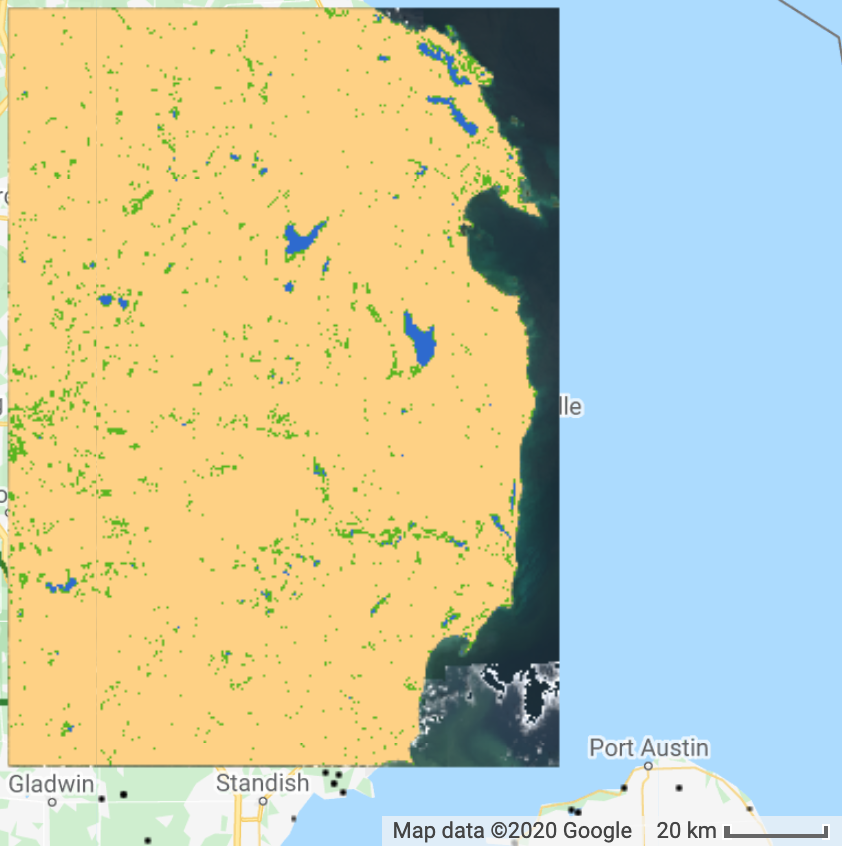


*Figure C5:* Variable Importance Graphs for the Lake Michigan Basin Classification



Appendix D

*Figure D1:* Classification Map for the Michigan Test Site with DSWE



*Figure D2:* Classification Map for the Michigan Test Site with Elevation

