Development of Automated Methods for Using Satellite Imagery to Monitor Changes in Vegetative Communities

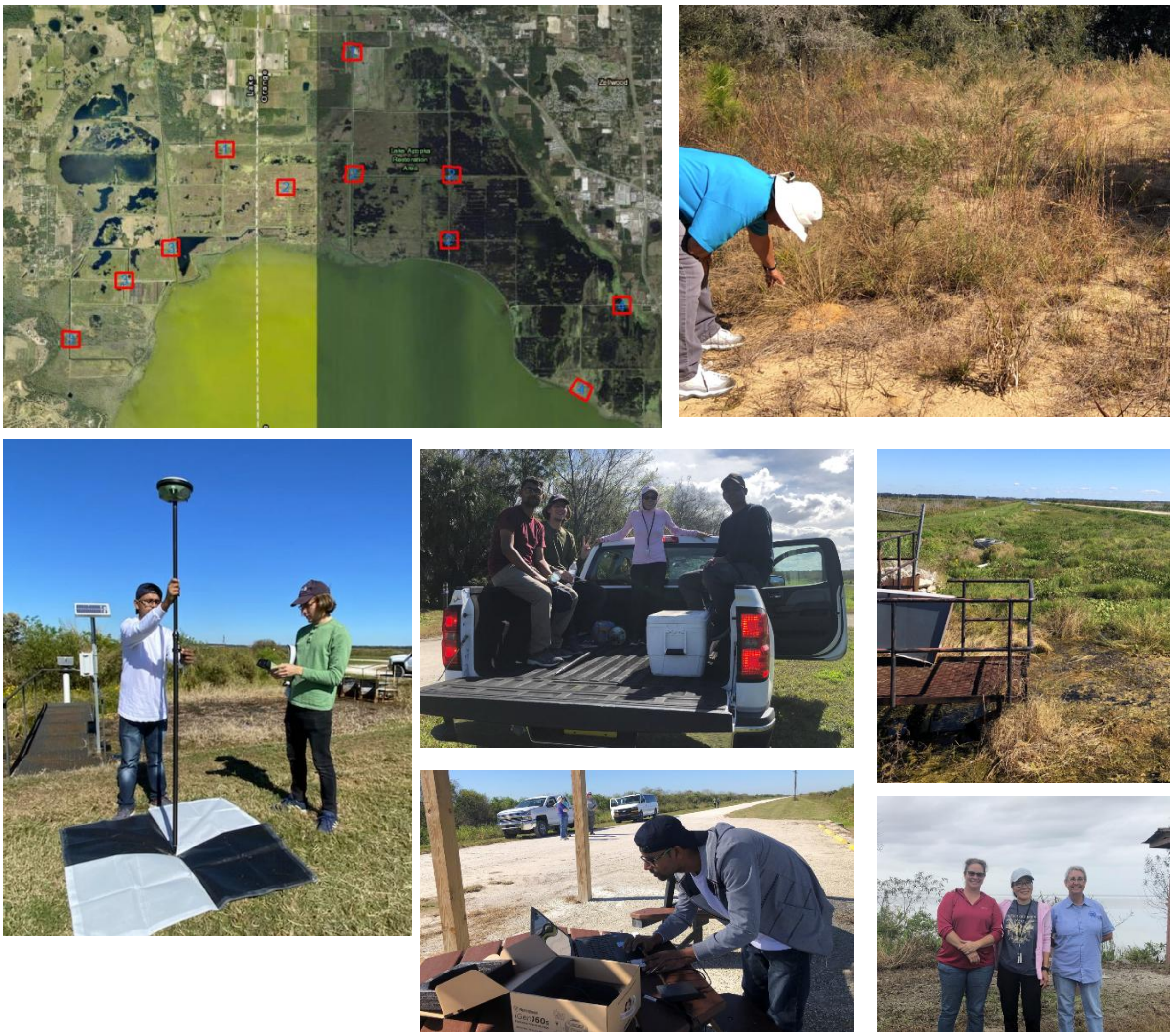
Report for Quarter 1/Task 5 in 2021-2022

Submitted to: St. Johns River Water Management District

Caiyun Zhang, PhD

Department of Geosciences, Florida Atlantic University

777 Glades Road, Boca Raton, FL 33431, USA, Tel: 561-297-2648; Email: czhang3@fau.edu



Collecting drone and field data in the Lake Apopka North Shore (LANS) during 11/15/2021-11/18/2021. Upper left: selected sites for field data collection; Upper right: Dr. Dianne Hall at District showing the identification of scrub community; Bottom left: PhD students Atiqur Rahman and David Brodylo setting up a ground target for drone data collection; Bottom middle: FAU field crew on District's truck and a student setting up laptop for drone data transfer; Bottom right: mixed herbaceous marsh in one field sector, and project PI Dr. Caiyun Zhang with District personnel Jodi Slater and Dianne Hall.

Contents

In this quarter, we mapped Lake Apopka North Shore Area (LANS) using the newly collected 2021 WorldView-2 (WV-2) imagery. LiDAR DEM was also included in the classification because LiDAR DEM improves classification accuracy as reported in previous quarters. We used eCognition to conduct image segmentation because the District has purchased the eCognition software package. We also collected drone data and field ground-truth data in this quarter during 11/15/2021-11/18/2021. In this quarter, we provided the field plan, collected field data, mapping accuracy, and 2021 vegetation maps for LANS.

1. Field Data Collection

1.1 Field plan

To effectively classify the newly collected 2021 WV-2 imagery, field ground-truth data are needed to use as training and testing samples in the classification, given the large change in plant communities occurred between 2017 and 2021 due to flooding from Hurricane Irma. After an online meeting and discussion with field specialist Jodi Slater and project manager Dr. Dianne Hall at the District, we selected 11 sites for drone and field data collection. In the selection, we considered the effects of eagle nest proximity (drone must stay 1000 feet away from nests), site accessibility by the District's truck and FAU's 11-seat van, suitability for ground target setup for drone data collection, distribution of sites across the project area, vegetation communities, capacity of the drone batteries, safety, and the logistics of planned field time schedule (time and cost budget). The selected sites are displayed in Figure 1, including 2 upland areas and 9 wetland areas. We made a plant community inventory which includes pictures taken by District personnel and text descriptions of each community to train FAU field crew to identify and collect field data. This inventory is included in the report as Appendix 1. We made a detailed agenda and communicated with District personnel about where, when, and which sites would be visited each day.

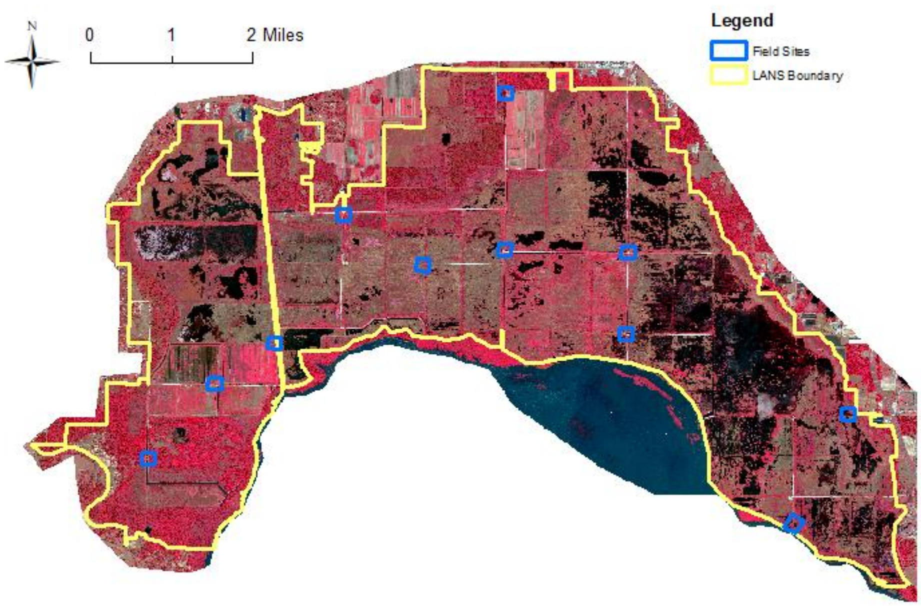


Figure 1. Selected sites (2 upland and 9 wetland areas) at LANS for field and drone data collection.

1.2 Field data collection

We collected field data using the ArcGIS Collector App. We first installed ArcGIS Collector App on our phone, and then set up point data collection in ArcGIS online (an ArcGIS online account must be created first). We created a new field for the point layer in ArcGIS online to save the community's name. To avoid error in typing in the field, a list of all community names was created and saved for the newly created field. In this way, we just needed to select the name and did not need to type in the field to avoid typing error and save collecting time. We created a group in ArcGIS online so all group members (PI and 3 PhD students) could access the same point layer and all other shared layers in the field. All collected points from different people could be synchronized to the same point layer. A tutorial of ArcGIS Collector can be found at https://www.youtube.com/watch?v=RJHHaEHdHBE. However, ESRI has discontinued support for Collector and ArcGIS now uses Field Maps to collect field data, which is similar to Collector. We divided our FAU crew members into two groups with one group (Atiqur Rahman and David Brodylo) mainly responsible for drone data collection and the other group (Caiyun Zhang and Mizanur Rahman) mainly responsible for ground-truth data collection. District Environmental Scientist Jodi Slater took the drone group to the target sites, and Environmental Scientist Dianne Hall led the field data collection group walking along the roads to identify plant communities. We added points and took pictures in the field. We used a DJI P4M system to collect drone imagery. This will be detailed in next quarter's report. Figures 2-13 show examples of our field work on the LANS.

Figure 2. Drone data collection group (David Brodylo and Atiqur Rahman) setting up a ground control point.

Figure 3. Drone data collection group (David Brodylo and Atiqur Rahman) setting up the DJI P4M drone in the field.

Figure 4. District Environmental Scientist Dr. Dianne Hall and FAU PhD student Mizanur Rahman collecting field data along a road. On the left is cattail, and to the right is shrub swamp.

Figure 5. A LANS observation tower was used to assist with field data collection due to a better view. We climbed two towers in the field trip.

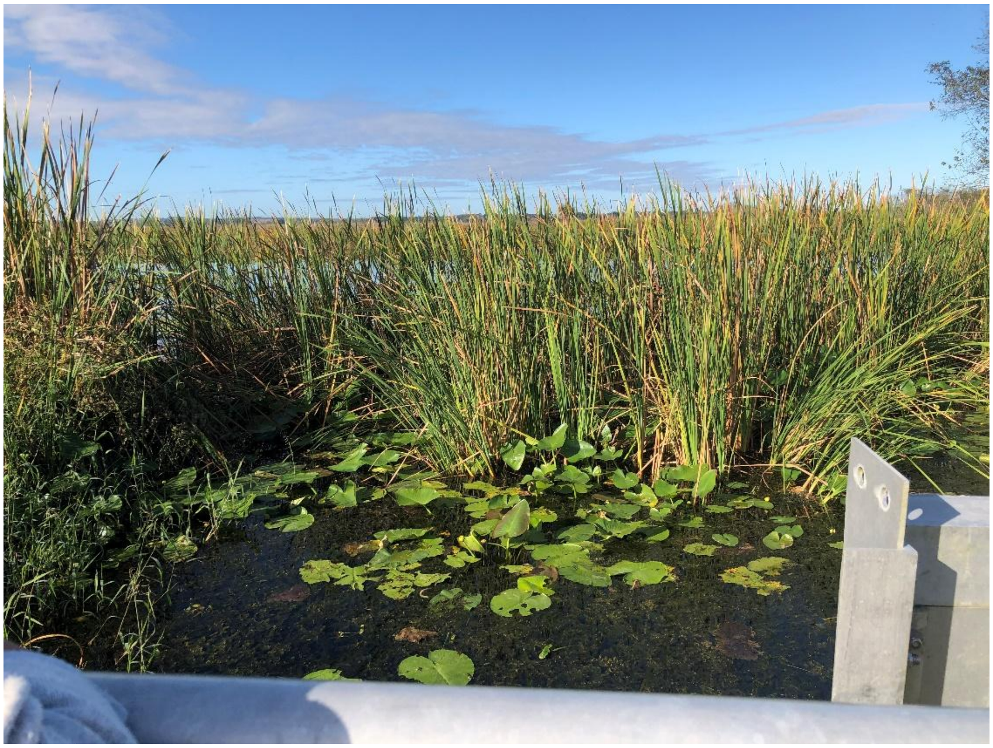


Figure 6. Observed Submerged Aquatic Beds (SAB) and Deep Marsh habitat (with spatterdock) in foreground with cattail in the background.

Figure 7. A pond covered by floating vegetation and bordered by shrub swamp in background and grass/sedge marsh in the foreground.

Figure 8. Observed shrub swamp in the field.

Figure 9. Observed pine plantation in the field.

Figure 10. Observed mixed herbaceous marsh and open water in the foreground and upland hardwood in the far distance as viewed from one of the observation towers.

Figure 11. Observed cattail in the field.

Figure 12. District Environmental Scientists Dr. Dianne Hall and Jodi Slater talking about the ecology and restoration of LANS to students.

Figure 13. FAU field crew moving to the next site.

After the field trip, we exported all collected field data from ArcGIS online and loaded it into ArcMap. In total, we collected 1495 points in the field. We manually checked and edited each point by aligning them to the 2021 WV-2 imagery. Using the field points of known plant communities enabled us to collect more imagery-based reference samples with similar characteristics like color, shape, and texture which are known as image interpretation keys in remote sensing. Guided by the field points, we collected more points off the imagery using ArcMap in office. We obtained a total of 7798 points including field ground-truth (1495) and imagery-based reference (6303) samples. These points were spatially joined to segments to identify segments/objects to be used in the training and testing procedure. We used the *k*-fold cross validation approach to train and test the classification by setting *k* to 10, i.e., 10% of the data were used to test the classification, and 90% were used to train the classifier, and this procedure iterated for 10 times. Figure 14 shows the location of the field and imagery based training/testing points.

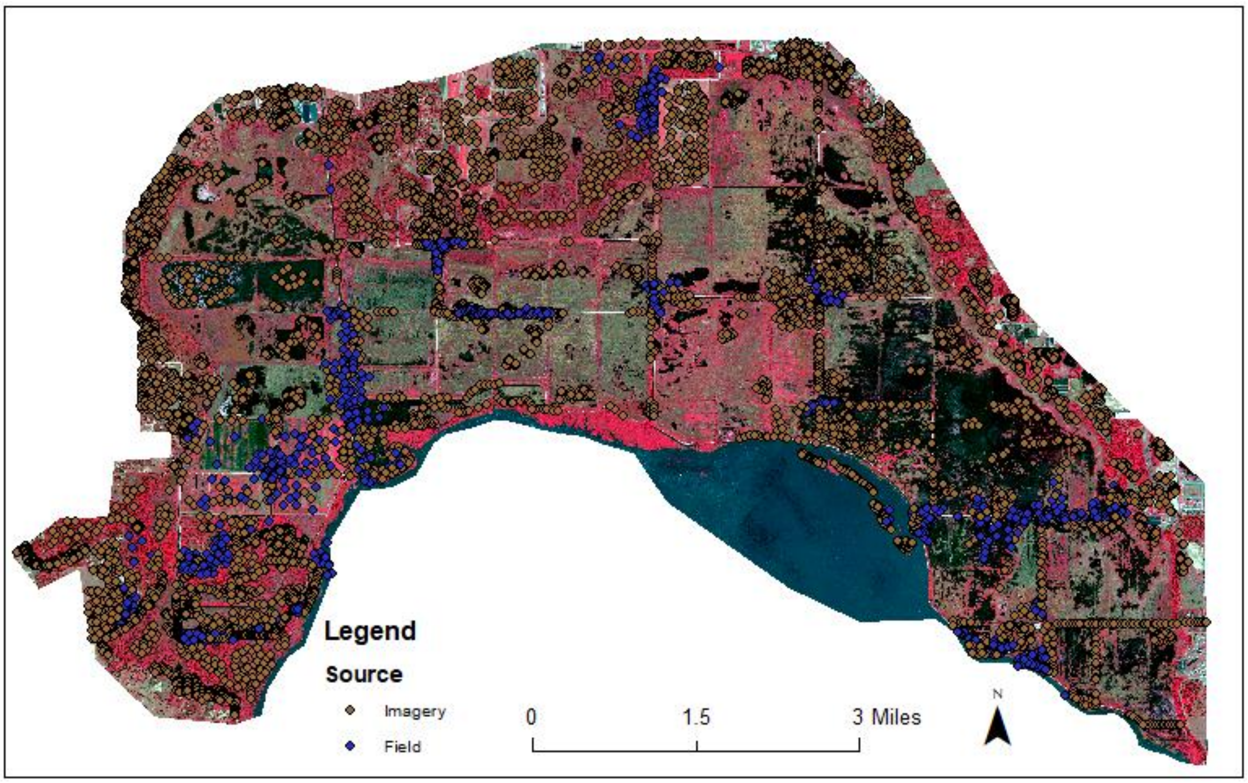


Figure 14. Point samples used to identify training and testing segments/objects in the classification.

2. Classification and Mapping

2.1 Segmentation result

The 32-bit WV-2 LANS imagery that the District provided was huge with a size of 73.3GB. Loading such a big imagery file into eCognition or ArcMap is challenging. Therefore, we converted the 32-bit imagery into unsigned 16-bit imagery using Arc Catalog. In Arc Catalog, right click on the imagery you want to convert and select “Export”, then select “Raster to

Different Format", where you can set the Pixel Type, and convert the 32-bit into 16-bit. This greatly reduced the imagery size to 14GB. We then loaded this 16-bit imagery into eCognition and ran the segmentation. The segmentation results from eCognition are provided in Figure 15. There were 252,099 segments created. The LANS is a large area. The 2017 WV-2 imagery that was used in the previous mapping of the LANS was provided in 2 parts and we ran the classification for the eastern and western areas separately. These smaller areas were easier to process in ArcMap and eCognition. For project areas larger than the 2021 LANS WV-2 imagery, we recommend splitting the imagery into smaller areas which are easier to handle in the software packages. Spectral and texture features were exported from eCognition into ArcGIS for training data processing and preparation.

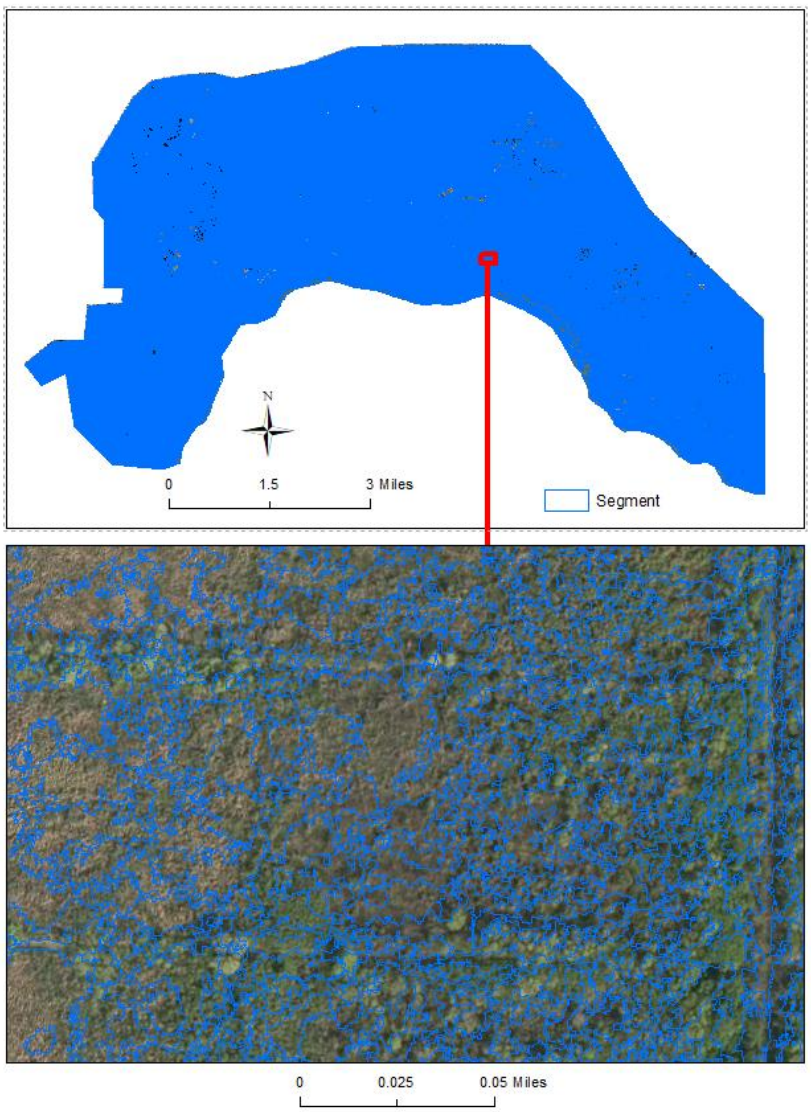


Figure 15. Segmentation from eCognition with a total of 252,009 segments for LANS.

2.2 Model performance

We trained the model in two different ways: using the imagery and field sample points combined and then using only the imagery derived points. The performance of Artificial Neural Network (ANN), Support Vector Machine (SVM), Random Forest (RF), and *k* Nearest Neighbor (kNN) using 8-band WV-2 imagery is displayed in Table 1. RF and SVM achieved a comparable accuracy with an Overall Accuracy (OA) above 70% and Kappa value of 0.68 or higher, regardless of which dataset was used to train the model. kNN generated the lowest accuracy using either method and was excluded from the ensemble analyses. Ensemble analysis of ANN, SVM and RF using the field and imagery sample points produced a slightly increased OA of 72.5% and Kappa value of 0.7 compared to any of the models by themselves.

When we trained each model using reference samples collected only from the imagery, the accuracy actually improved (Table 1). This is because when we picked training samples from the imagery, we chose points with a clear shape, color, and other distinct features on imagery as reference samples. We didn't select points as reference samples if we weren't sure what they were. This led to the models only being trained on the "best" points. However, in the field, we put points in the communities that we saw, even though these patches/regions did not have distinct signature features on the imagery. Field visits can help label imagery reference samples more accurately, and field samples are more realistic. However, field data can be only collected over accessible/visible areas, and field data collection is labor intensive. Therefore, only limited field data were obtained in LANS. A combination of field sampling and image-based reference data collection is ideal for this project. The ground-truth data were used in both the training and testing procedures via the *k*-fold cross validation, which preserved the independence of the training and testing sample points. To conduct a supervised digital image classification like the machine learning classifiers used in this project, training data are mandatory. Training data can be collected from imagery (requires prior knowledge of the project area and what the community looks like on imagery) or from field visits by developing signature keys.

Table 1. Performance of each machine learning classifier and ensemble analysis for LANS; EA is ensemble analysis of ANN, SVM, and RF.

Using reference samples collected from field and imagery					
	ANN	kNN	SVM	RF	EA
OA (%)	67.2	62.1	70.3	71.4	72.5
Kappa Value	0.64	0.59	0.68	0.69	0.70
Using reference samples collected from imagery only					
OA (%)	72.7	66.2	75.3	75.2	77.0
Kappa Value	0.70	0.63	0.73	0.73	0.75

2.3 Classified maps

Classification maps from machine learning algorithms, and ensemble analysis are displayed in Figure 16. In general, all classifiers produced a similar spatial pattern. Pine Plantation (PI), Pine Flatwoods (PF), and Cattail (CT) were delineated in the 2021 classified map. These communities were not mapped in the 2017 reference map. Both PI and PF were grouped into Upland in the 2017 reference map. PI was observed in the field and training samples were collected, while PF training samples were collected on imagery based on the state Land Use Land Cover (LULC) map with a LCLU code of 4410 (Coniferous Pine). We did not see PF in the field, but it is present in the Upland. We observed large patches of Cattail in the field and collected corresponding training samples. Both the 2017 reference map and 2017 classified maps did not delineate Cattail. Large patches of Submerged Aquatic Vegetation (*Hydrilla*) were observed at F cell of the West Marsh, which were not present in the 2017 imagery (Figure 17). The presence of *Hydrilla* was confirmed by aerial imagery collected by the District in May of 2021 and shown in Figure 17. We collected image-based reference samples for *Hydrilla*, and it subsequently mapped well in the F cell areas of the LANS. For most regions, three classifiers (ANN, SVM and RF) generated consistent classifications, as shown in green in the uncertainty map in Figure 16 (f). Red regions show a "warning sign" where the classification from each classifier was completely inconsistent with each other and can be used as an indicator of where further training points should be gathered.

2.4 Analysis

The error matrix from the models that used the training set comprised of both imagery and field data showed that classification of Hydric Hammock (HH), Herbaceous Mixed Marsh (HM), Wet prairie/wet pasture (WP), and Grass/Sedge (GS) had an accuracy less than 50%. HH was confused with Upland Hardwood (UH), as illustrated in the earlier 2017 image classification. The use of the LiDAR DEM did not enable the models to differentiate between forested communities at different topographic elevations. For the 2017 satellite imagery, we did not map Cattail (CT) because it wasn't a designated community on the aerial imagery-derived maps provided by the District. However, during the field visit we saw many patches of CT, and therefore we mapped CT for the 2021 imagery. HM was mainly confused with CT, which is understandable, since cattail can be a component of HM. HM was also classified as Shrub Swamp (SS); small shrubs can be a component of HM communities as well. In the field we saw many patches of a mixture of HM and SS, and thus this is understandable given the lack of a hard boundary between the two communities. We saw limited WP in the field and WP was confused with SS based on the error matrix. The classification accuracy of GS was very low (6%) in the ensemble model, with the greatest accuracy in the RF model (20%). GS was classified as levee (levee with grass can be confused with GS), CT, and HM. This also is understandable given that CT is grass-like and both grasses and sedges are components of herbaceous marshes. We had very limited training samples of GS. In the field we did not see many homogeneous GS areas. Maybe GS can be eliminated in the classification. We saw limited Willow Swamp, Scrub, Cypress and did not see Palmetto Prairie which appeared as a minor community in the aerial-based 2017 reference map. We did not map these communities due to limited training samples. The error matrix is large and uploaded as a separate Excel file. We generated two error matrices using a combination of field and image-based reference samples as training samples, and then using image-based training samples alone. In the next quarter, we will use the field and drone collected data to assess the accuracy of the classified maps and create corresponding error matrix. Field and drone data will not be used in the training procedure and used in testing only as an independent dataset.

3. Files for Quarter 1, 2021-2022

Products in this quarter include ArcGIS shapefiles of ANN, SVM, RF, Ensemble Analysis classification, Uncertainty Analysis; two error matrixes in Excel format; field point samples in ArcGIS shapefile; and this report. They are uploaded to the District's ftp site.

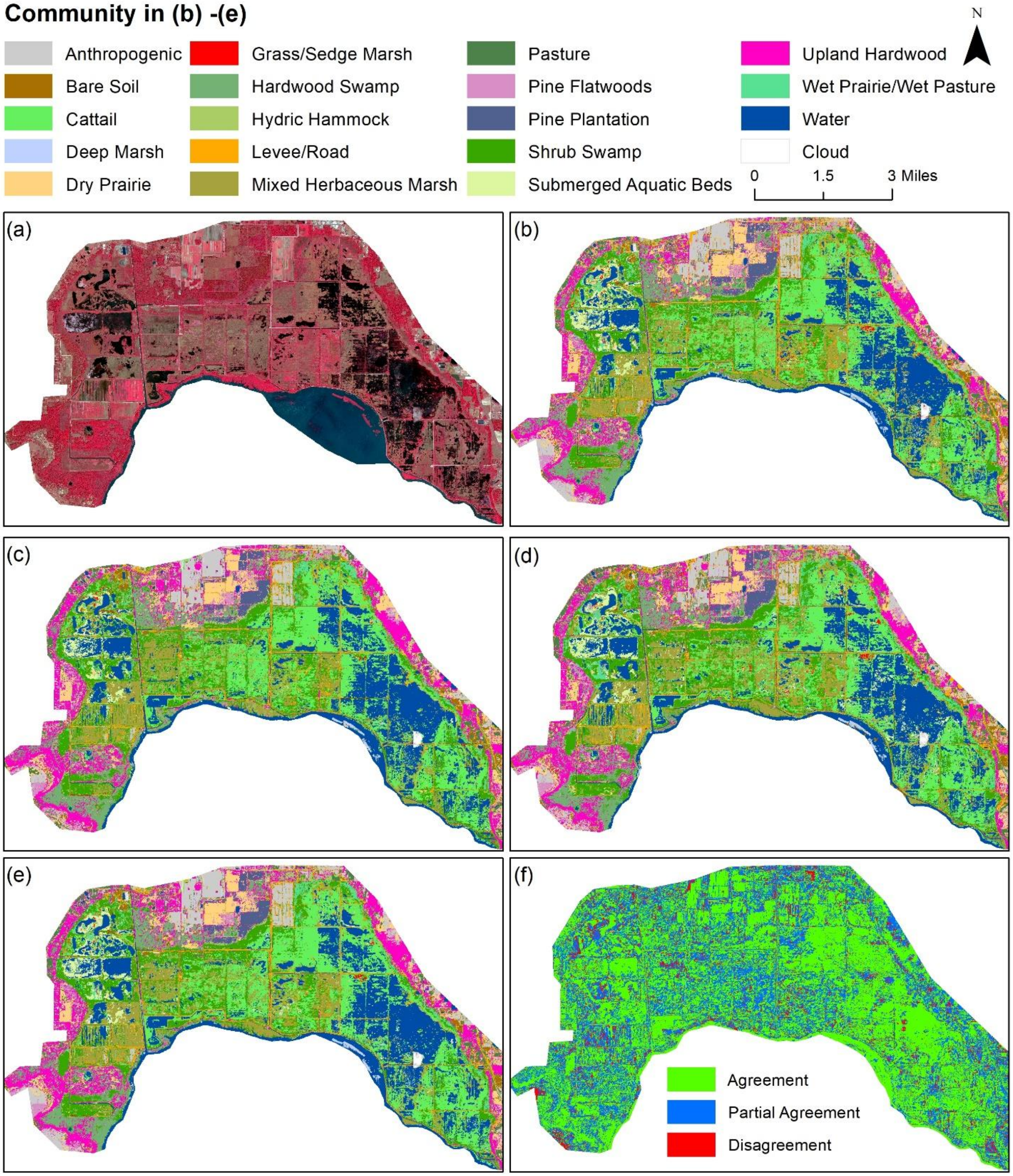


Figure 16. (a) 2021 WV-2 pansharpened imagery; Object-based classifications for 2021 imagery from (b) SVM, (c) RF, (d) ANN, and (e) Ensemble analysis. (f) Uncertainty from ANN, SVM, and RF.

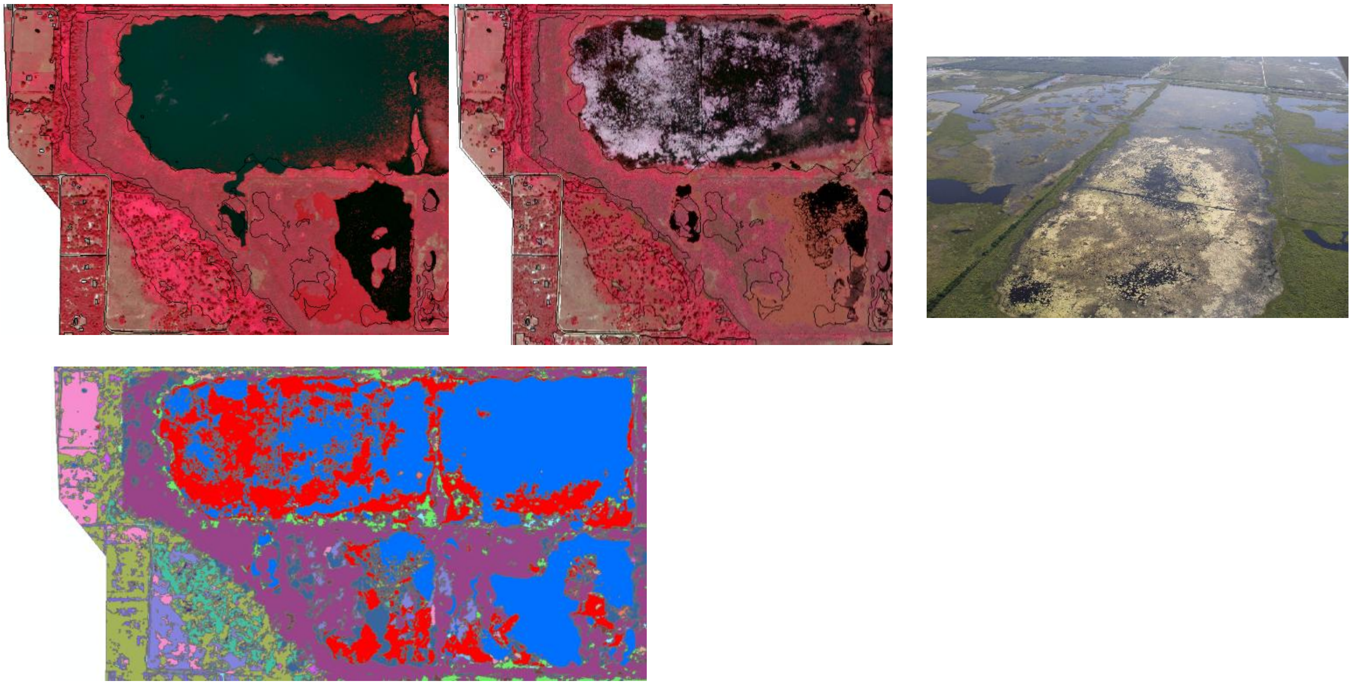


Figure 17. F cell in west marsh area of the LANS in 2017 WV-2 imagery (upper left), 2021 WV-2 imagery (upper middle), May 2021 aerial view (upper right), and identified SAB *Hydrilla* highlighted in red in the 2021 classified map (bottom).

Appendix 1 Community Inventory Used in the Field

To effectively conduct field data collection, the District provided us the pictures of each community and we made this community inventory to train the field crews.

1, AN, Anthropogenic: Areas of agricultural land (including orchards, groves, and row crops, but NOT pasture or pine plantations), and buildings, parking lots, spoil piles, development, etc. Includes bare areas associated with anthropogenic disturbance.


AN-Crop


AN-Spoil

2, BS, Bare Soil: Open areas with > 70% cover of bare soil with no vegetation (e.g., shoreline, beach, mudflats, barrens); not associated with anthropogenic disturbance or salt flats.

3, DM, Deep Marsh: Areas containing ≥ 40% cover of bottom-rooted species with floating leaves (e.g., *Nymphaea, Nuphar, Nelumbo, Brasenia, Nymphoides*) or deep -water emergent species (e.g., *Schoenoplectus*); may also include *Utricularia* spp and may occur as a littoral zone around lakes.

4, DP, Dry Prairie: Areas containing > 50% cover of mixed upland grasses and other herbaceous species (e.g., *Aristida, Andropogon, Xyris, Rhexia*) with < 50% cover of low shrubs (e.g., *Serenoa, Ilex glabra, Lyonia, Vaccinium, Quercus minima*) and few trees.

Foreground is dry prairie, and background is pine flatwoods.

5, HH, Hydric Hammock: Areas containing ≥ 70% canopy cover of *Quercus laurifolia, Sabal palmetto, Magnolia virginiana, Juniperus, Celtis, Liquidamber*, and/or *Ulmus spp.* and having shorter hydroperiods than Hardwood Swamp; tree canopy may commonly include *Pinus*, but will rarely include *Taxodium* spp.; understory may include *Serenoa repens*, *Callicarpa americana* and *Psychotria nervosa* or ferns

6, HM, Mixed Herbaceous Marsh: Areas containing a mixture of broadleaf emergents, grasses, sedges, and other spp. (e.g., *Pontederia, Sagittaria, Hydrocotyle, Persicaria);* small shrubs may be included as a small part of the community (e.g., *Cephalanthus, Kosteletzkya, Hibiscus*).

7, HS, Hardwood Swamp: Areas containing ≥ 70% mixed wetland tree species (e.g., *Acer, Taxodium, Fraxinus, Nyssa, Ulmus, Annona*), cypress may be a significant component, but < 70%; includes swamps and communities lying in the floodplain of rivers and streams; understory typically doesn't include *Serenoa repens* (saw palmetto).

8, LR, Levee/Road: Paved or unpaved roads or levees with grassed, gravel, limerock or dirt roads at their apex; levees without roads should be classified based on the vegetation they support.

9, OW, Water: Areas such as lakes, impoundments, rivers, canals, ditches, and other areas of open water which may support free-floating plant species.

The first picture also shows cattail in the background.

10, PA, Pasture: Areas similar in composition to Dry Prairie, but with recent evidence of pasture/cattle management (e.g., fencing, feeding or water troughs) and the presence of exotic forage grasses (*Paspalum notatum, Hemarthria, Panicum repens, Cynodon*).

11, SS, Shrub Swamp: Areas containing ≥ 70% cover of mixed shrub species (e.g., *Salix*, *Morella*, *Baccharis*, *Sambucus*, *Ludwigia*, *Cephalanthus*, *Kosteletzkya*, *Hibiscus*); may include small trees (e.g., *Acer*, *Gordonia*, *Persea*, *Ilex*, *Sabal*, *Schinus*).

12, UH, Upland Hardwood: Areas containing a mixture of hardwood species with pines < 40% and palms < 70%; may be dominated by mesic oak species.

13, WP, Wet Prairie/Wet Pasture: Areas containing a mixture of grasses, sedges, rushes, and forbs typically classified as facultative-wet or facultative; long-hydroperiod species should not be present; category should also be used for re-flooded pastures and wet, unimproved pastures.

14, SAB, Submerged Aquatic Beds: Areas containing rooted aquatic plants with photosynthetic tissues below the water surface.

Sawgrass also appears in the top picture.

15, PP, Palmetto Prairie: Areas containing > 50% cover of *Serenoa repens* and low shrubs with few to no trees; < 50% cover of mixed upland grasses and herbs.

16, PF, Pine Flatwoods: Areas typically containing > 40% cover of *Pinus* spp. with an understory of low shrubs, grasses and forbs; where intensive burns have occurred, trees may be defoliated and understory vegetation may dominate.

Below two pictures are pine plantations: plant neatly shows in rows

18, CY, Cypress: Areas containing ≥ 70% cover of bald or pond cypress; includes cypress domes, swamps, strands and lakeshore variants.

19, FD, Forested Flatwood Depressions: Areas containing mixed communities of *Taxodium*, *Pinus*, *Sabal*, bays or deciduous hardwoods in shallow depressions; may be located within areas of mesic hardwoods or pine flatwoods.

No Picture, we did not see this in LANS field trip.

20, CP, Cabbage Palm Hammock: Areas similar in species composition to Hydric Hammocks, but containing ≥ 70% cover of *Sabal palmetto*

21, CT, Cattail: Areas containing ≥ 70% cover of *Typha* spp.

22, GS, Grass/Sedge Marsh: Areas containing ≥ 70% cover of obligate or facultative-wet grass or *sedge spp*.; other non-graminoid species and small shrubs may be present.

23, WS, Willow Swamp: Areas containing ≥ 70% cover of *Salix caroliniana*

We saw sparse Willow Swamp in the field.

24, SC, Scrub: areas containing ≥ 70% cover of scrub species and scrub oak species, with or without an overstory of *Pinus clausa*; areas of bare white sand may be visible and should be included as part of the community.

We saw limited scrub in the field.

25, SP, Spartina: areas containing ≥ 70% cover of *Spartina bakeri*.

We did not see SP in the field.