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

Final Report for Contract 35547-1

Submitted to: St. Johns River Water Management District

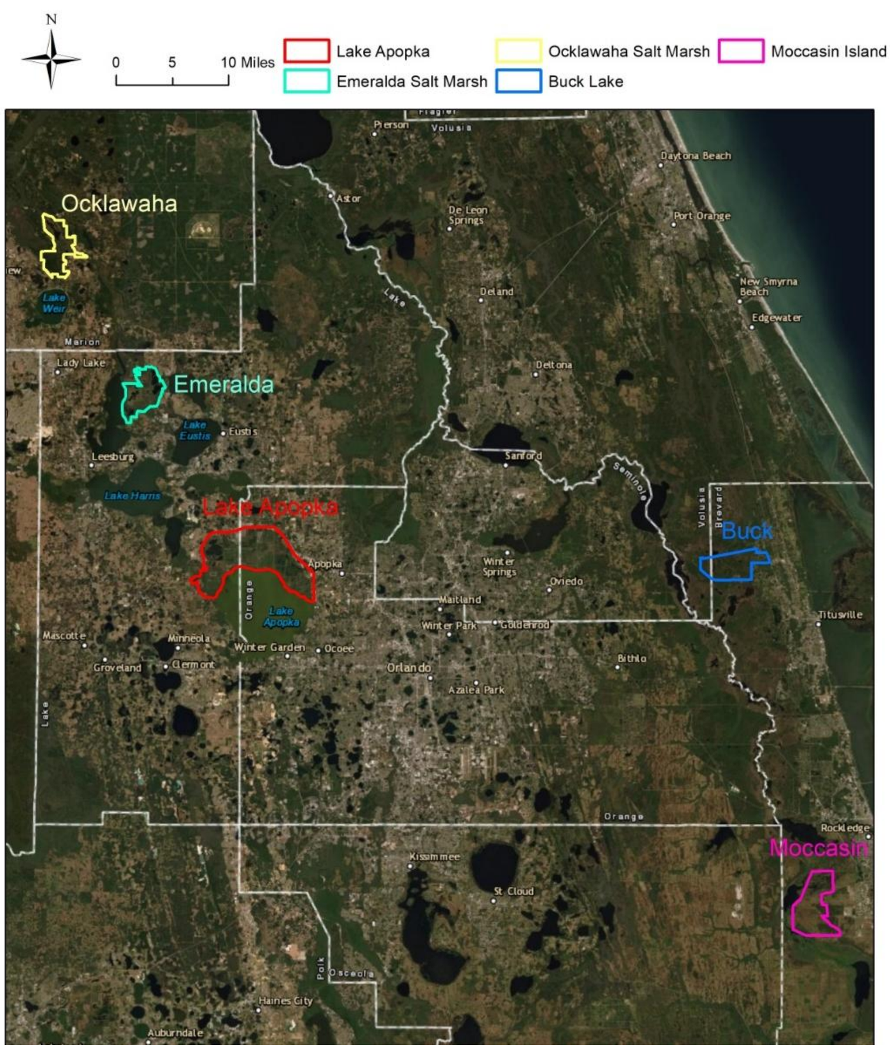
By

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



Five mapped project areas covering 55,955 acres.

Contents

During an online meeting with Dr. Dianne Hall on 16 August 2022 (meeting summary is in Appendix I), we discussed five components to include in the final report to wrap up this project: 1) a summary of project year 1 results with a focus on comparing aerial versus satellite data sources, selection of software and mapping methodology and problems that might arise with the change to machine learning and satellite imagery; 2) a summary of project year 2 results with a focus on software package selection, tutorial development and methods to improve overall mapping accuracy; 3) discussion of issues and solutions in the mapping procedure; and 4) other concerns relevant to wetland community mapping not specified in the Scope of Work. These components are provided below.

1. Mapping wetland communities

The District has been managing more than 430,000 acres of wetlands which need to be mapped regularly to monitor effects of management and environmental changes on habitats. The District has been collecting digital aerial imagery with four spectral bands (Red, Green, Blue, and Near-Infrared (NIR)) from airborne sensors to generate vegetation maps through a manual interpretation procedure. It takes months to collect digital aerial imagery, including time spent on flight planning, sensor selection, and final aerial imagery generation and rectification. After the acquisition of aerial imagery, it can take several months to generate a vegetation map product through the manual interpretation procedure, especially when mapping large areas. This conventional vegetation mapping procedure is time-consuming and labor-intensive and can result in vegetation maps that are inaccurate due to the length of time between imagery acquisition and ground truthing of the mapped environments.

Satellite remote sensing has been frequently cited as a cost-effective and labor-saving technique for monitoring and mapping wetlands. The operational WorldView-2/3 (WV-2/3) satellites are the most technologically advanced high-resolution optical sensors, collecting 46-centimeter (cm) panchromatic and 1.85-meter (m) multispectral imagery. WV-2/3 sensors have a comparable spatial resolution to aerial imagery, but they add four extra spectral bands: coastal, yellow, red edge and an additional NIR band to the traditional four (Red, Green, Blue, NIR) bands. The satellite revisit period of WV-2/3 (~2 days) in conjunction with their large spatial coverage make spaceborne sensors more appealing for monitoring wetland response to changing environments than using airborne sensors, especially over large areas.

Contemporary machine learning based digital imagery classification techniques have been widely used as an alternative to manual interpretation. A combination of semi-automated/automated machine learning with WV-2/3 satellite imagery accelerates the wetland mapping process, making it more effective for informing decision making in wetland management than the traditional manual interpretation of aerial imagery.

As a first step in this project, we designed a machine learning ensemble protocol for mapping over 20 wetland plant communities across five project areas covering 55,955 acres using 8-band 0.3-m pansharpened WV-2 imagery collected in 2017 and 2019. The areas that were mapped included Ocklawaha Prairie Restoration Area (OPRA) (6,450 acres), Emerald Marsh Area (7,090 acres), Moccasin Island Area (9,850 acres), Buck Lake Area (7,260 acres), and Lake Apopka North Shore (LANS) (25,305 acres). These maps, generated through machine learning and using satellite imagery, were compared to 2017 maps of these same areas that were produced using aerial imagery and manual interpretation. This comparison demonstrated that the machine learning procedure was much more time efficient and could generate a reasonable overall accuracy (OA) based on reference samples collected from the original manually interpreted 2017 maps, as summarized in Table 1. It took only a few weeks to finish one project area, rather than months. The machine learning procedure can also utilize other data sources such as LiDAR DEMs (digital elevation models) and NDVI (normalized difference vegetation index) indices in the mapping procedure.

Table 1. Summary of the overall accuracy (OA) in percent of the machine learning mapping procedure for the five project areas mapped using pansharpened 8-band WV-2 imagery in this contract.

	2017						2021
	OPRA	Emerald	Moccasin Island	Buck Lake	West LANS	East LANS	LANS
OA	80.3	84.4	84.8	71.2	77.5	80.0	72.5

Though testing results are encouraging, there are two major constraints using the satellite-based machine learning mapping procedure: WV-2/3 imagery product selection and machine learning data processing. Cloud cover is a major issue in using satellite imagery for mapping in Florida, as well as other tropical and subtropical areas, especially during the rainy season. Less than 10% cloud cover should be stipulated for tasked imagery acquisition of WV-2/3 imagery. Imagery may have to be obtained during the dry season, which may cause additional problems if ground truthing is done during the wet season. In the machine learning mapping procedure, training sample selection and specification of required input parameters of each machine learning algorithm will impact mapping results. Training samples are mainly collected from the WV-2 imagery using an imagery interpretation procedure based upon the historical maps. The accuracy of the old maps will impact the labeling of training samples which will consequently have effects on the new map generation. However, ground surveys of areas with heterogeneous habitats or habitats that are similar to one another can mitigate this problem. In the machine learning training procedure, a trial-and-error procedure can be used to identify optimal parameters for classifiers. Minor communities with a small coverage in the project areas might be difficult to map due to limited training samples and subsequent confusion with other communities. These minor communities should be dropped. Also note that the image segmentation procedure may take more than 10 hours for object generation and exporting of relevant features, depending on the scale set during the multiresolution segmentation procedure.

2. Tutorial development

A major focus of this contract was to develop detailed tutorials for using the designed machine learning ensemble mapping protocol. In machine learning, selection of software packages is critical. The first step is to select a segmentation software that can effectively generate image objects from the 8-band WV-2 imagery in order to conduct object-based vegetation mapping. If fine resolution imagery products (<5 m) are used, object-based image analysis (OBIA) is the state-of-art technique for mapping. We attempted to use open-source software for this step. We tested two open-source image segmentation packages including TerraView, and Large Scale Mean Shift in OTB/QGIS. We also assessed the Mean Shift segmentation in ArcMap, and Edge segmentation in ENVI. However, none of these segmentation algorithms could produce a comparable segmentation result to the multiresolution segmentation algorithm integrated in Trimble's eCognition software. For example, TerraView would only segment part of the input imagery, and QTB/QGIS and Edge segmentation in ENVI couldn't identify small patches. eCognition not only produces ideal segmentation results at several scales, but also can export valuable features of the imagery, such as texture measurements, to improve classification accuracy. Consequently, the commercial eCognition Developer 10.0 for image segmentation was purchased by the District in order to conduct object-based vegetation mapping.

The second step is to select appropriate machine learning and classification packages. We selected the open-source WEKA and R packages for machine learning training and classification, and the commercial package ArcMap 10.2, which the District already owns, for training sample selection and final vegetation mapping. WEKA does not require any coding skills and contains many of the most successful machine learning functions such as Support Vector Machine (SVM), Random Forest (RF), and Artificial Neural Network (ANN). We just need to click buttons and set required parameters to complete the training and classification in WEKA. We also developed R scripts as an alternative to WEKA. This requires coding skills but is more automated and can generate formatted error matrices, classifications, McNemar test results, and ensemble classifications. WEKA does not have the capability to combine the outputs of several classification models into an ensemble function, nor can it calculate the McNemar test to determine which classifications are significantly different from one another. Thus, we developed corresponding R scripts to use outputs from WEKA to complete the ensemble classification and McNemar test, for those who would rather use the WEKA software to select appropriate classifiers. Consequently, the tutorial provides two separate pathways for generating a plant community map from an ensemble analysis of the best classifiers.

3. Discussion

Once the WV-2/3 imagery is acquired, a major step impacting the mapping procedure is the training sample selection. Ideally, training samples should be selected based on a stratified sample selection strategy through field visits, but in practice field sample collection is difficult especially in inaccessible areas. Misalignment between field samples and imagery can cause issues too. The WV-2/3 imagery products have a very fine resolution (0.3 m in this project), which allows reference sample collection by a visual interpretation procedure based upon old maps. However, you must be careful not to perpetuate errors from the old map in the new one. A combination of this visual interpretation with samples collected by field visits is more effective than the application of imagery-based or field-based sample collection alone. Special care should be paid in obtaining training samples from communities containing similar groups of species. For example, mixed herbaceous marsh communities, may contain cattail and small shrubs, which will be present in the shrub swamp community, and cattail communities as well. Training points should be selected over homogeneous areas where only one community is present. For instance, training samples for cattails should be taken in areas where cattail is at least 70% of the community. Likewise, training points for shrub swamp should only be taken from areas where shrubs are at least 70% of the community. Confusing labeling of some communities based on older maps may also add uncertainties in the mapping results. For example, older District maps and Florida Land cover/Land use (LCLU) shapefiles don't use the same community definitions as those that are currently being used and the communities have to be matched to one another or some communities may need to be split or combined. In one instance, we found pine flatwoods, pine plantation, upland hardwood, and cabbage palm hammock communities all labeled differently in the LCLU maps and these communities are only labeled as Uplands in older District maps. Thus, field visits are required to correctly label training samples for communities like these. Minor communities (e.g., less than 25 acres in coverage) are difficult to map due to limited training samples and these communities should be dropped. This requires prior knowledge/information of the project area.

4. Others

All the required tasks specified in the Scope of Work have been completed, with the exception of final revisions to the tutorials and the Final Report. The District expressed some other concerns relevant to wetland community mapping. The District is interested in using the vector polygon vegetation map products generated under this contract for change analysis (i.e., is the change analysis methodology different using these new maps), and also in using these procedures to map a single specific community type, if the other communities are not of interest (e.g., mapping just *Phragmites australis*). Our plant community map products are dissolved polygons with the plant communities labeled. The products can be utilized directly in ArcMap for change analysis either at the plant community level or the community type level (e.g., Herbaceous Wetland, Shrub Upland, etc.). Although it was not required by this contract, Appendix II provides a detailed tutorial for change analysis using our map products in ArcMap. The main step is to utilize the Intersect function in ArcMap which can break down overlapping areas of two maps into smaller areas and then perform the change analysis on each smaller area.

To identify only a specific community, a similar classification procedure can be followed, making sure that the training samples that are selected are as indicative of that community as possible. There are binary coding techniques (i.e., a classification procedure to classify imagery into only target and non-target groups) in remote sensing, but in general they use the same logic of classification, and all require training samples. The unsupervised procedure might be more effective for this binary type of classification. Unsupervised classifiers cluster the imagery into a number of clusters specified by a user based on their spectral similarities. Unsupervised classification does not require the training sample selection procedure, but it does require that the final clusters be identified. WEKA has integrated some unsupervised clustering functions such as K-Means Clustering (a tutorial in Youtube: https://www.youtube.com/watch?v=c5Qi5P2s-Ng). The Arff file to be used for final prediction (see the Mapping Tutorial V_12.docx) can be used for unsupervised clustering in WEKA following the youtube tutorial, and then we can label each cluster in ArcMap into defined classifications. For binary coding, users can simply label the target, such as the Cattail community, corresponding to one or more clusters (multiple clusters can be Cattail), and label all the other clusters as “others”.

5. Deliverables

This quarter’s deliverables included the updated mapping tutorial (Mapping Tutorial_V12.docx) that includes the R script to utilize the WEKA outputs for running the McNemar test, and constructing the ensemble analysis (EnsembleAnalysis_McNemarTest _V2.r). The entire tutorial was written using only the data from the West Marsh of the Lake Apopka North Shore to be consistent throughout. Also included were the testing datasets for the R script of (EnsembleAnalysis_McNemarTest _V2.r), as well as the testing datasets and instructions for the change analysis (Appendix II) and this Final Report.

Appendix I: Discussion of What to Include in the Final Report

Attendants: Dianne Hall, and Caiyun Zhang

Time: 8/16/2022, 10:00-11:30

Form: Microsoft Team Meeting online

Summary:

In this online meeting, we talked about potential further collaboration for next phase mapping project using satellite imagery and techniques we developed in this project. We focused on items to be included in the final report in Quarter 4, 2021-2022. Based on the meeting, the following components should be included in the final report.

1) A component focused on Project Year 1 results: discuss the advantages using spaceborne satellite WorldView-2 imagery for wetland mapping compared to the previously used 4-band digital aerial imagery; discuss the advantages using semi-automated/automated modern machine learning techniques for wetland mapping compared to the previously used manual interpretation; discuss the fast, cost-efficient procedure we developed in the project and its demonstration in five project areas.
2) A component focused on Project Year 2 tutorial development: discuss the selection of open-source software packages vs. commercial software packages, WEKE, R script, eCognition, and other tested packages in project year 2, and final identification of the software packages in the project.
3) A component discussing issues and potential solutions in the mapping procedure: examples include minor community mapping, training sample selection, training samples for some specific community such as cattail.
4) Revising the tutorials: will be delivered as separate files, including drop kNN in McNemar test in the R script, discussion why drop kNN based on WEKE outputs, R script for running machine learning, R script for testing using field samples or drone reference samples.
5) Discussion of potential solutions for change analysis based on the map products generated in the project, solution for identification of a specific community and etc. which are not included in the current contract but is needed in District.

We may need to revise the EA mapping dataset based on comments from other staff. This will not be included in the final report.

Appendix II: Change Analysis in ArcMap

Caiyun Zhang, Mizanur Rahman

8/29/2022

This tutorial instructs how to conduct change analysis based on our map product, a dissolved vector polygon shapefile with Plant Community attribute. Examples are provided based on a subset of map products of Lake Apopka North Shore for years 2021 and 2017.

- Create Community Type attribute in ArcMap. District is interested in Community Type level change, we thus should create a text type field to save the Community Type, for example we named it as Class21, and Class17 for the 2021 and 2017 map, respectively. We then assign the Community Type for each polygon by aggregating the Plant Community into Community Type based on the Appendix 2 of the Scope of Work which identifies the Plant Community Classifications and Definitions, the results are shown in Figure 1 in ArcMap. The area of each Community Type can be added and calculated using Calculate Geometry function.

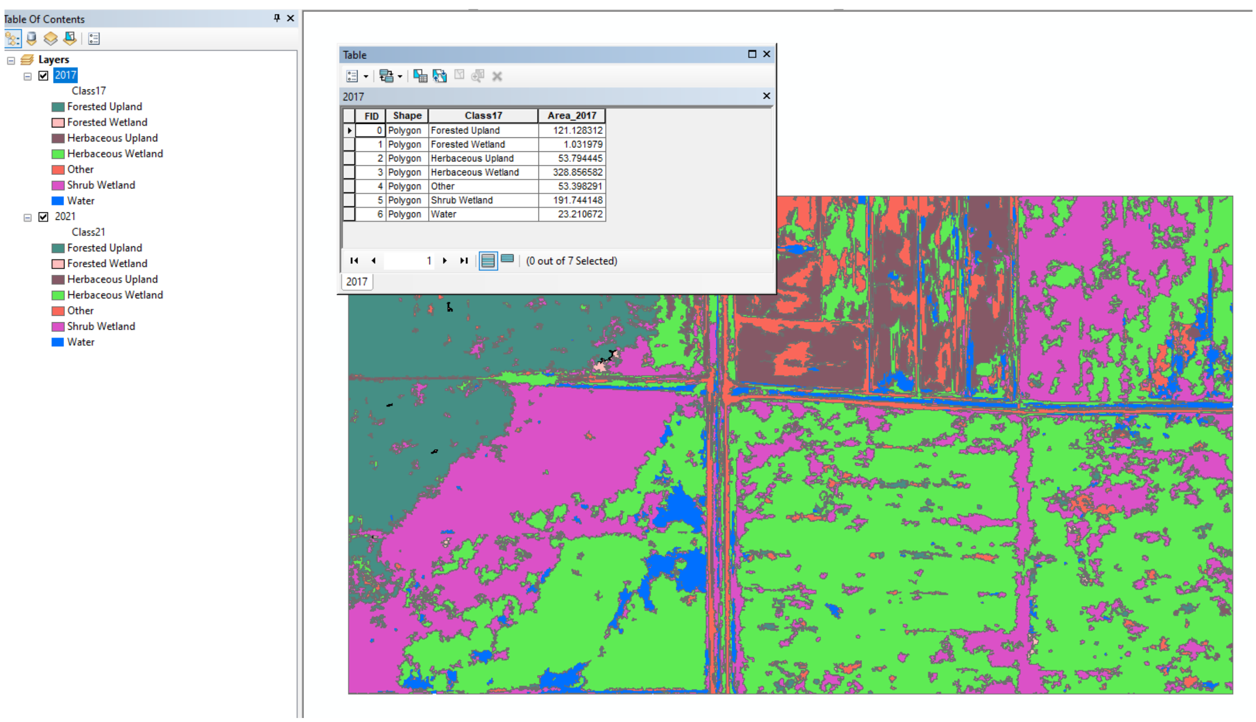


Figure 1. Subset of 2017 and 2021 plant community type map.

- Go to Geoprocessing->Intersect, and save it as change17_21.shp, as shown in Figure 2.

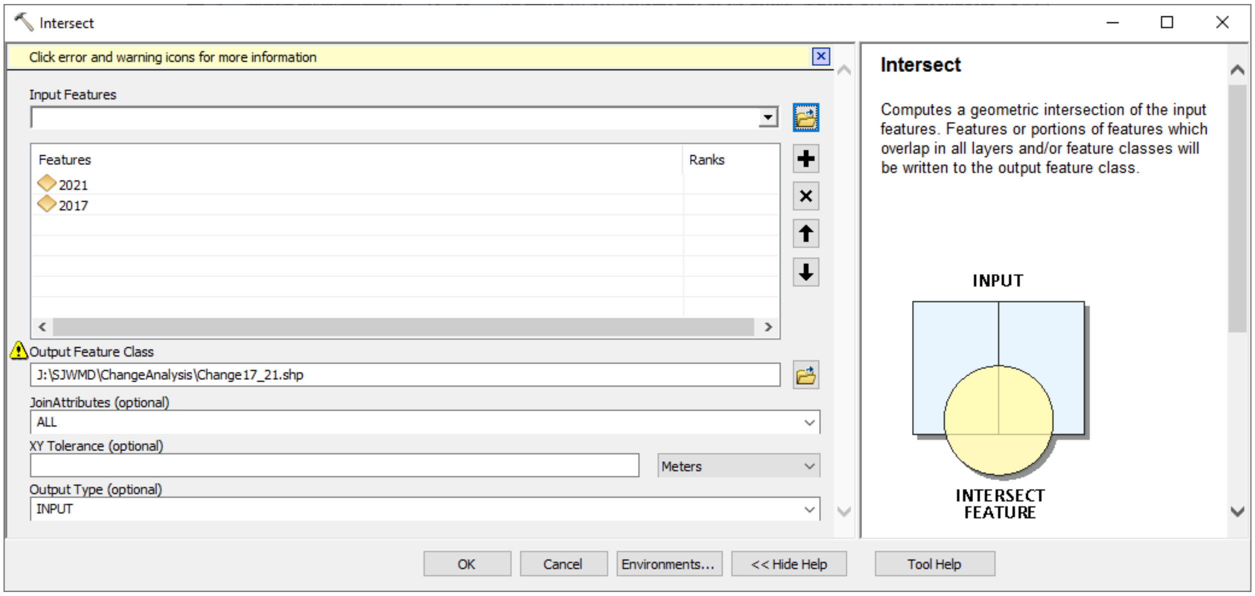
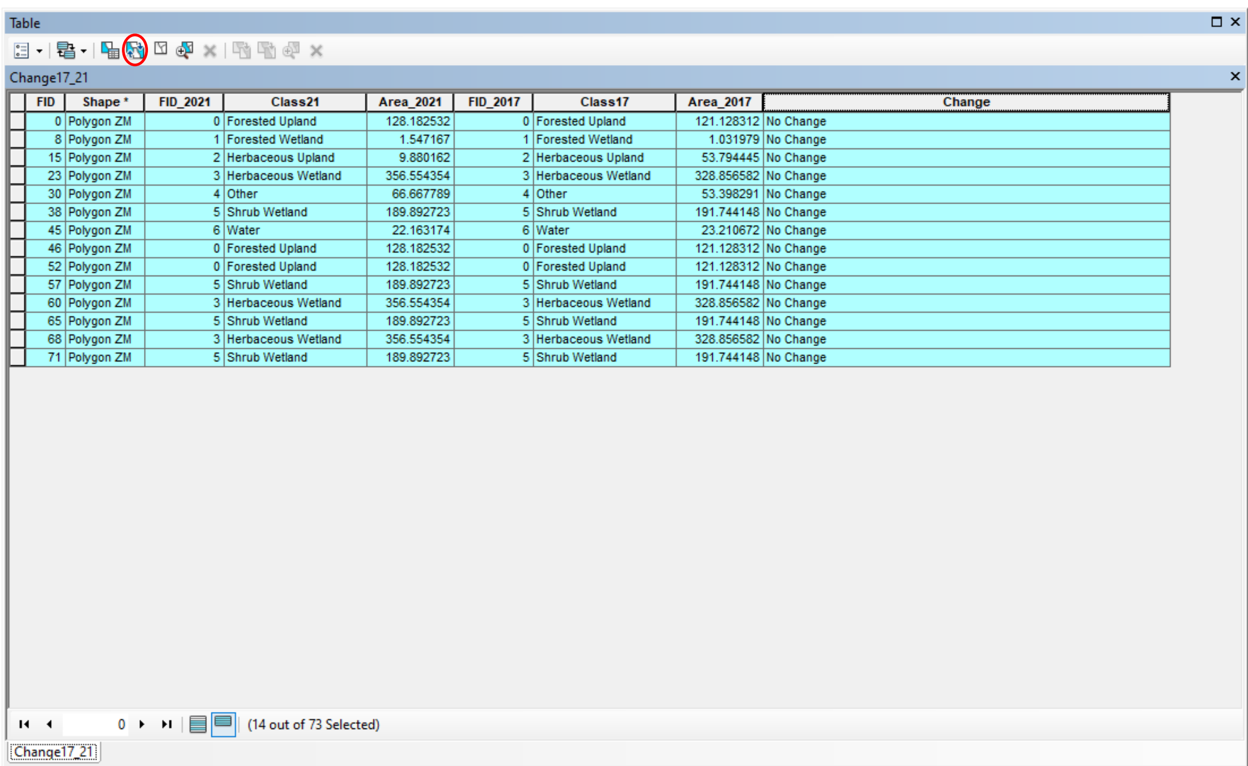


Figure 2. Intersect of 2021 and 2017.

- Create a new text field named as Change
- In the Attribute Table of Change17_21, select Class21=Class17, and then assign “No Change” to the Change attribute, as shown in Figure 3.



FID	Shape *	FID_2021	Class21	Area_2021	FID_2017	Class17	Area_2017	Change
0	Polygon ZM	0	Forested Upland	128.182532	0	Forested Upland	121.128312	No Change
8	Polygon ZM	1	Forested Wetland	1.547167	1	Forested Wetland	1.031979	No Change
15	Polygon ZM	2	Herbaceous Upland	9.880162	2	Herbaceous Upland	53.794445	No Change
23	Polygon ZM	3	Herbaceous Wetland	356.554354	3	Herbaceous Wetland	328.856582	No Change
30	Polygon ZM	4	Other	66.667789	4	Other	53.398291	No Change
38	Polygon ZM	5	Shrub Wetland	189.892723	5	Shrub Wetland	191.744148	No Change
45	Polygon ZM	6	Water	22.163174	6	Water	23.210672	No Change
46	Polygon ZM	0	Forested Upland	128.182532	0	Forested Upland	121.128312	No Change
52	Polygon ZM	0	Forested Upland	128.182532	0	Forested Upland	121.128312	No Change
57	Polygon ZM	5	Shrub Wetland	189.892723	5	Shrub Wetland	191.744148	No Change
60	Polygon ZM	3	Herbaceous Wetland	356.554354	3	Herbaceous Wetland	328.856582	No Change
65	Polygon ZM	5	Shrub Wetland	189.892723	5	Shrub Wetland	191.744148	No Change
68	Polygon ZM	3	Herbaceous Wetland	356.554354	3	Herbaceous Wetland	328.856582	No Change
71	Polygon ZM	5	Shrub Wetland	189.892723	5	Shrub Wetland	191.744148	No Change

Figure 3. Assign “No Change” to unchanged polygons/records.

- Switch record to changed polygons by click the Switch Selection button, highlighted in Figure 3.
- Using Field Calculator function to assign Change= [Class17]+"-"+ [Class21]. note here both Class17 and Class 21 are text type, as shown in Figure 4.

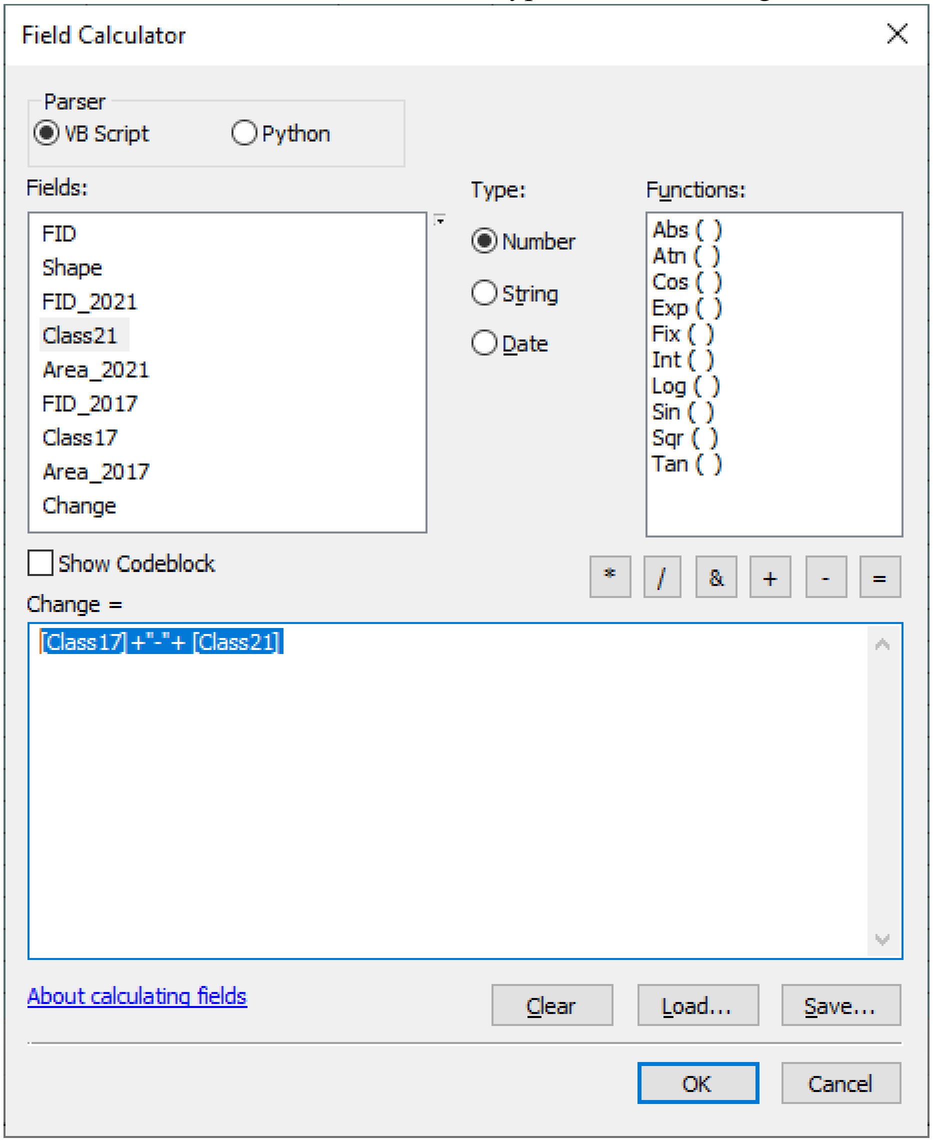


Figure 4. Assign change label (from what to what) to changed polygons/records.

The final attribute table of the Change17_21 is shown in Figure 5.

Table

Change17_21

FID	Shape *	FID_2021	Class21	Area_2021	FID_2017	Class17	Area_2017	Change
7	Polygon ZM	1	Forested Wetland	1.547167	0	Forested Upland	121.128312	Forested Upland-Forested Wetland
14	Polygon ZM	2	Herbaceous Upland	9.880162	0	Forested Upland	121.128312	Forested Upland-Herbaceous Upland
20	Polygon ZM	3	Herbaceous Wetland	356.554354	0	Forested Upland	121.128312	Forested Upland-Herbaceous Wetland
47	Polygon ZM	3	Herbaceous Wetland	356.554354	0	Forested Upland	121.128312	Forested Upland-Herbaceous Wetland
27	Polygon ZM	4	Other	66.667789	0	Forested Upland	121.128312	Forested Upland-Other
33	Polygon ZM	5	Shrub Wetland	189.892723	0	Forested Upland	121.128312	Forested Upland-Shrub Wetland
53	Polygon ZM	5	Shrub Wetland	189.892723	0	Forested Upland	121.128312	Forested Upland-Shrub Wetland
40	Polygon ZM	6	Water	22.163174	0	Forested Upland	121.128312	Forested Upland-Water
1	Polygon ZM	0	Forested Upland	128.182532	1	Forested Wetland	1.031979	Forested Wetland-Forested Upland
21	Polygon ZM	3	Herbaceous Wetland	356.554354	1	Forested Wetland	1.031979	Forested Wetland-Herbaceous Wetland
34	Polygon ZM	5	Shrub Wetland	189.892723	1	Forested Wetland	1.031979	Forested Wetland-Shrub Wetland
2	Polygon ZM	0	Forested Upland	128.182532	2	Herbaceous Upland	53.794445	Herbaceous Upland-Forested Upland
48	Polygon ZM	0	Forested Upland	128.182532	2	Herbaceous Upland	53.794445	Herbaceous Upland-Forested Upland
9	Polygon ZM	1	Forested Wetland	1.547167	2	Herbaceous Upland	53.794445	Herbaceous Upland-Forested Wetland
22	Polygon ZM	3	Herbaceous Wetland	356.554354	2	Herbaceous Upland	53.794445	Herbaceous Upland-Herbaceous Wetland
49	Polygon ZM	3	Herbaceous Wetland	356.554354	2	Herbaceous Upland	53.794445	Herbaceous Upland-Herbaceous Wetland
58	Polygon ZM	3	Herbaceous Wetland	356.554354	2	Herbaceous Upland	53.794445	Herbaceous Upland-Herbaceous Wetland
28	Polygon ZM	4	Other	66.667789	2	Herbaceous Upland	53.794445	Herbaceous Upland-Other
35	Polygon ZM	5	Shrub Wetland	189.892723	2	Herbaceous Upland	53.794445	Herbaceous Upland-Shrub Wetland
59	Polygon ZM	5	Shrub Wetland	189.892723	2	Herbaceous Upland	53.794445	Herbaceous Upland-Shrub Wetland
41	Polygon ZM	6	Water	22.163174	2	Herbaceous Upland	53.794445	Herbaceous Upland-Water
3	Polygon ZM	0	Forested Upland	128.182532	3	Herbaceous Wetland	328.856582	Herbaceous Wetland-Forested Upland
54	Polygon ZM	0	Forested Upland	128.182532	3	Herbaceous Wetland	328.856582	Herbaceous Wetland-Forested Upland
10	Polygon ZM	1	Forested Wetland	1.547167	3	Herbaceous Wetland	328.856582	Herbaceous Wetland-Forested Wetland
16	Polygon ZM	2	Herbaceous Upland	9.880162	3	Herbaceous Wetland	328.856582	Herbaceous Wetland-Herbaceous Upland
29	Polygon ZM	4	Other	66.667789	3	Herbaceous Wetland	328.856582	Herbaceous Wetland-Other
36	Polygon ZM	5	Shrub Wetland	189.892723	3	Herbaceous Wetland	328.856582	Herbaceous Wetland-Shrub Wetland
55	Polygon ZM	5	Shrub Wetland	189.892723	3	Herbaceous Wetland	328.856582	Herbaceous Wetland-Shrub Wetland
61	Polygon ZM	5	Shrub Wetland	189.892723	3	Herbaceous Wetland	328.856582	Herbaceous Wetland-Shrub Wetland
42	Polygon ZM	6	Water	22.163174	3	Herbaceous Wetland	328.856582	Herbaceous Wetland-Water
69	Polygon ZM	6	Water	22.163174	3	Herbaceous Wetland	328.856582	Herbaceous Wetland-Water
0	Polygon ZM	0	Forested Upland	128.182532	0	Forested Upland	121.128312	No Change
8	Polygon ZM	1	Forested Wetland	1.547167	1	Forested Wetland	1.031979	No Change
15	Polygon ZM	2	Herbaceous Upland	9.880162	2	Herbaceous Upland	53.794445	No Change
23	Polygon ZM	3	Herbaceous Wetland	356.554354	3	Herbaceous Wetland	328.856582	No Change
30	Polygon ZM	4	Other	66.667789	4	Other	53.398291	No Change
38	Polygon ZM	5	Shrub Wetland	189.892723	5	Shrub Wetland	191.744148	No Change
45	Polygon ZM	6	Water	22.163174	6	Water	23.210672	No Change
46	Polygon ZM	0	Forested Upland	128.182532	0	Forested Upland	121.128312	No Change
52	Polygon ZM	0	Forested Upland	128.182532	0	Forested Upland	121.128312	No Change
57	Polygon ZM	5	Shrub Wetland	189.892723	5	Shrub Wetland	191.744148	No Change
60	Polygon ZM	3	Herbaceous Wetland	356.554354	3	Herbaceous Wetland	328.856582	No Change
65	Polygon ZM	5	Shrub Wetland	189.892723	5	Shrub Wetland	191.744148	No Change
68	Polygon ZM	3	Herbaceous Wetland	356.554354	3	Herbaceous Wetland	328.856582	No Change
71	Polygon ZM	5	Shrub Wetland	189.892723	5	Shrub Wetland	191.744148	No Change

0 (0 out of 73 Selected)

Change17_21

Figure 5. Labeled changed and unchanged polygons/records. This can then be mapped in ArcMap as shown in Figure 6.

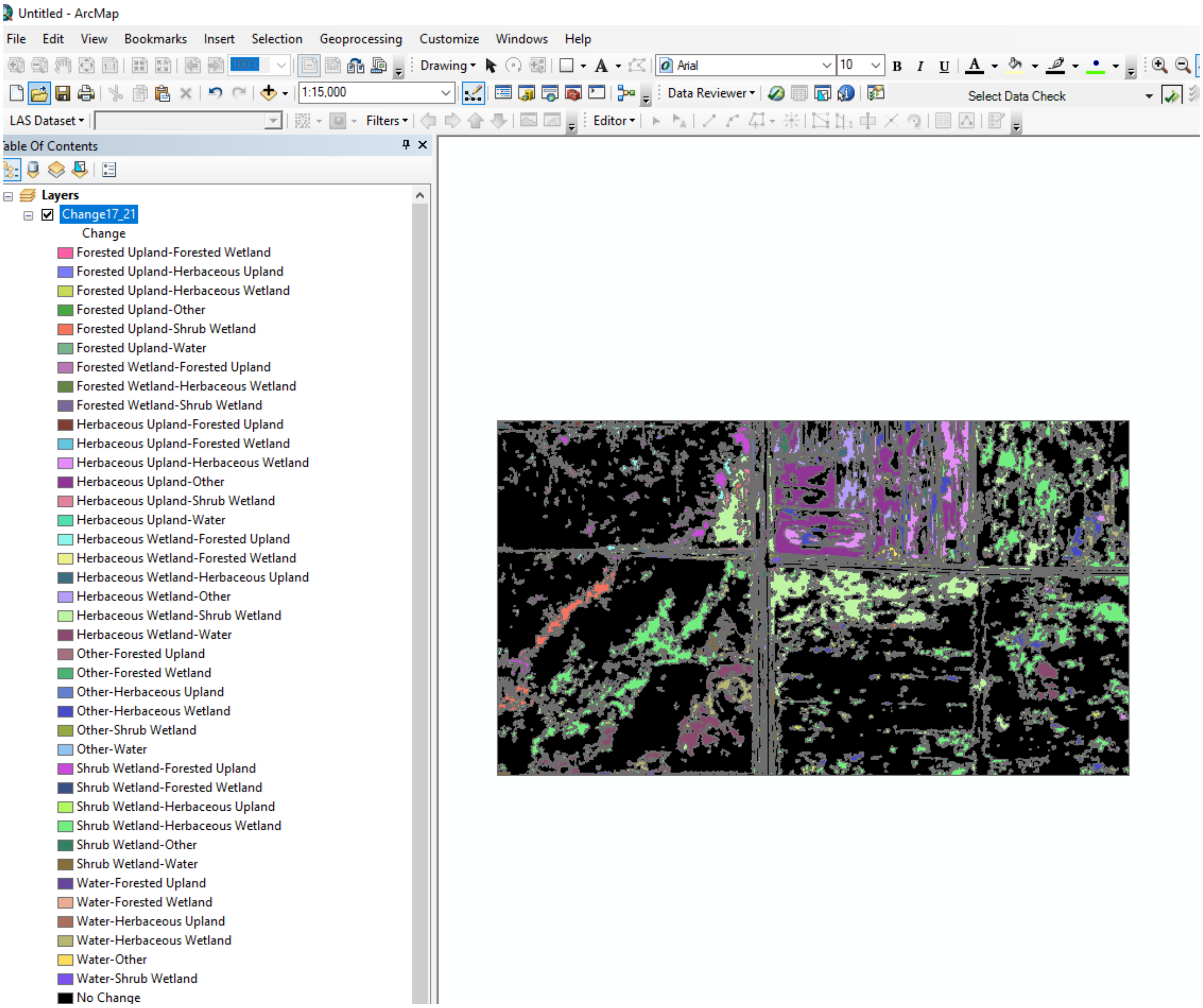


Figure 6. Changed and unchanged area mapping in ArcMap.