Note to the reviewer: To obtain the training data, contact kelly.streich@ct.gov

EELGRASS AREA CLASSIFICATION IN ARCGIS PRO

Final Draft September 25, 2024

CT Department of Energy and
Environmental Protection
79 Elm Street
Hartford, CT 06106

The University of Vermont
Spatial Analysis Laboratory 205
George D. Aiken Center
Burlington, VT 05405

Project funded by Connecticut Department of Energy and Environmental Protection through an EPA Long Island Sound Study Assistance Grant.

Contents

Table of Contents

1.0 Project Background

The Connecticut Department of Energy and Environmental Protection (CT DEEP), with funding assistance from the Long Island Sound Study pursued a pilot project to develop an eelgrass assessment methodology for embayments in Long Island Sound. Eelgrass (*Zostera marina*), the main species of seagrass found in Long Island Sound, provides habitat for fish species and other invertebrates and has long been recognized as a signal of a healthy embayment. This project explores a methodology using Unmanned Aircraft System (UAS) technology to collect the imagery of eelgrass beds and automated feature extraction using spatial technology to determine the aerial extent of eelgrass.

To initiate the project, CT DEEP developed a Quality Assurance Project Plan (*Beebe Cove Pilot Project – Developing Methods to Use an Unmanned Aircraft System (UAS)/Drone for Coastal Embayment Assessments*; RFA# 21090, approved 6/23/2022). Following approval of the QAPP, eelgrass imagery was collected in Beebe Cove, located in Groton CT on 7/19, 7/20, 8/3 and 8/4 in 2022. The imagery was collected using a high-resolution 45 MP true-color photogrammetry camera known as the DJI Zenmuse P1; a high resolution 10-band multispectral camera known as the Micasense RedEdge-MX Dual camera; the DJI Matrice 300 RTK drone; and a high-precision handheld RTK (Real-Time Kinematic) enabled GPS receiver known as the Emlid Reach RS2. The imagery was processed into 2D orthomosaic maps using a photogrammetric software known as PIX4Dfields.

Field data to ground truth the drone and spatial analyses were collected on 7/21/2022 by the University of Connecticut (UConn). This involved visiting fifty-five stations located in and around the area of known eelgrass in Beebe Cove and collecting camera images. The camera images were later evaluated in the laboratory for eelgrass presence and density.

Following processing and initial analyses of the data, CT DEEP contracted with the University of Vermont Spatial Analysis Lab (UVM-SAL) to explore various automated feature extraction approaches. CT DEEP and UVM-SAL held a number of meetings throughout development of the methodology. To document the process, the following sections provide a description of the evaluated classification methods, purpose of the document, image classification workflow, and a training exercise. Specifically, the training exercise (Section 5.0) provides user-friendly detailed instructions to complete an object-based classification methodology to delineate the area of eelgrass in Beebe Cove, conduct an accuracy assessment, and calculate the areal extent of eelgrass.

2.0 Evaluation of Classification Methods

To initiate this effort, features from the imagery were identified as eelgrass and non-eelgrass and classified to delineate the overall area of eelgrass. Data inputs sourced from Beebe Cove included both RGB and multispectral imagery (red, green, blue, and near-infrared spectral bands), and buffered sample points (reference data) indicating whether features were eelgrass or not. The imagery was used to create image objects/segments (groups of pixels), and sample points were then used to classify eelgrass features.

Following preliminary analyses, five approaches (Table 1) were identified as potential tools to delineate eelgrass from multispectral imagery using object-based classification methods in ArcGIS Pro. The order of accuracy is based on a qualitative, visual inspection and comparison of results. These approaches were applied to RGB imagery, but the process of image preparation was not executable in ArcGIS Pro and therefore, did not result in a final delineation of eelgrass. Additionally, the same five methods were also tested using pixel-based methods in ArcGIS Pro. Table 2 summarizes the pixel-based approaches. While pixel-based approaches may be optimal for some classifications and in some scenarios, it was found that the object-based approaches provided more optimal delineation of eelgrass for this project. As such, the object-based workflow and classification method as applied to multispectral imagery is the subject of this document.

Table 1 – Object-Based Classification Methods, Order of Accuracy, and Requirements.

Method	**Order of Accuracy**	**Requires Training Samples**	**Requires Segmented Image**
Maximum Likelihood	1	Yes	Yes
Random Trees	2	Yes	Yes
Support Vector Machine	3	Yes	Yes
K-Nearest Neighbor	4	Yes	Yes
ISO Cluster	5	No	Yes

Table 2 – Pixel-Based Classification Methods, Order of Accuracy, and Requirements.

Method	**Order of Accuracy**	**Requires Training Samples**	**Requires Segmented Image**
K-Nearest Neighbor	1	Yes	No
Support Vector Machine	2	Yes	No
Random Trees	3	Yes	No
Maximum Likelihood	4	Yes	No
ISO Cluster	5	No	No

3.0 Document Purpose

This document provides instructions to complete object-based classifications on raster imagery using ESRI's ArcGIS Pro classification tools. The training exercise is specifically tailored to the eelgrass delineation used to classify multispectral imagery from Beebe Cove. All object-based methods require that segmentation be applied to a raster or image prior to executing the classification tool. Four of the five classification methods (all except ISO Cluster) require training samples to evaluate the imagery and classify accordingly (see Table 1). Training samples may be based on ground-truthing field data or created by drawing polygons using analyst knowledge of the imagery and site conditions. To learn more about of segmentation and classification tools, please see ESRI's website: Understanding segmentation and classification, Classify.

The segmentation process as well as the classification methods listed in this document contain a variety of configurable parameters that affect the output products. For the initial testing, default values were used for all classification methods. Some parameter changes may improve classification outputs. However, it is recommended that these changes be evaluated and tested by the analyst, as the optimal configurations will depend on the project context and input datasets. Default settings will likely suffice in many cases (as with the initial analysis in this project), but that is also project-, context-, and dataset- dependent. Additional information on segmentation and classification parameters can be accessed on ESRI's website: Segmentation, Classify.

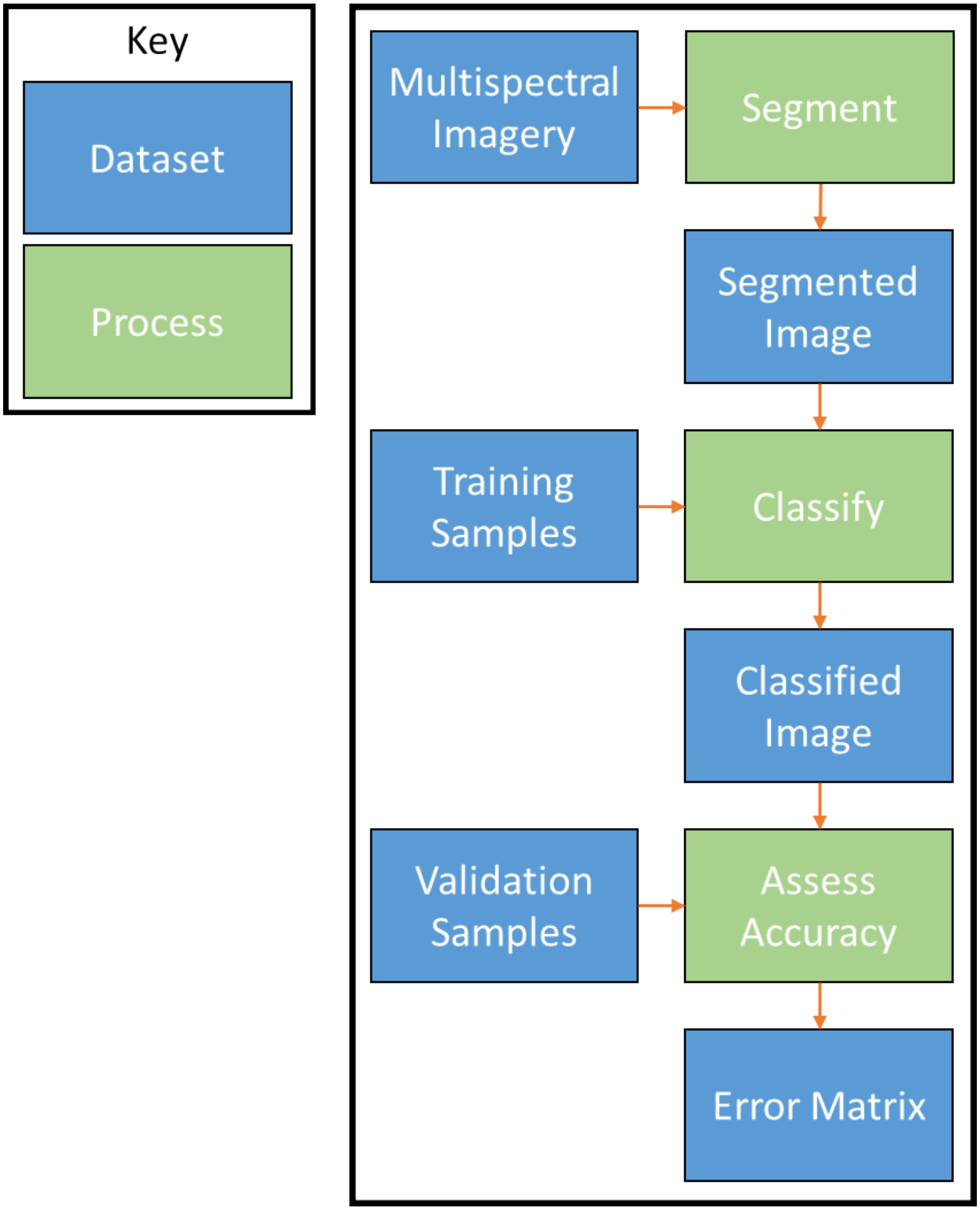
Time estimates are included after section headers where processing steps occur, to contextualize and approximate how long the analyst can expect each process to take. This will help the analyst budget time, and detect if a process is taking too long and therefore may be erroneous.

The steps in this document were tested on a Windows 10 machine, using ArcGIS Pro 3.1.1. Processing times will vary depending on the computer/workstation. For overall processing time, object-based classifications take longer to complete than the pixel-based classifications. The section headers that require processing contain an estimate of time (e.g., 30-60 minutes) within the section titles. These are just estimates. From testing, it appears one important factor in processing time efficiency is the number of parallel processes running on the workstation. Sufficient RAM is necessary to process the datasets, though it is not critical to processing efficiency in the same way the parallel processing is.

It should be noted that outputs may differ if a different version of ArcGIS Pro is used. In testing this workflow, differences in the segmentation and classification rasters were noticed when comparing results from version 3.1.1 to 3.0.3.

4.0 ArcGIS Image Classification Workflow

The workflow diagram illustrates the complete methodology applied to delineate the eelgrass area of Beebe Cove. The processing steps are provided in the following Training Exercise.



5.0 Training Exercise

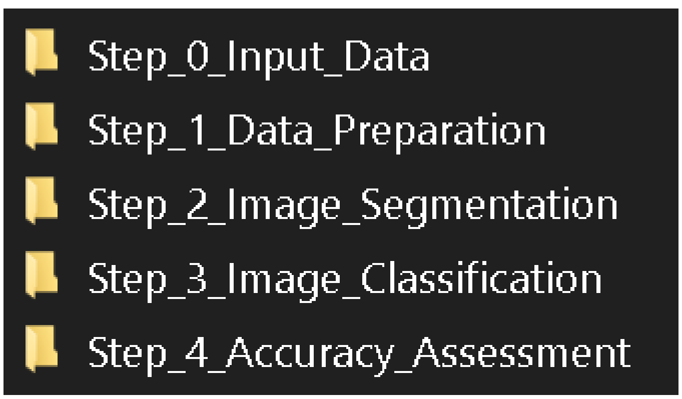
In this section, you will complete areal delineation of eelgrass using Beebe Cove multispectral imagery in ArcGIS Pro. The steps to delineation include preparing the training samples, image segmentation and classification. You will also conduct an accuracy assessment and calculate the delineated area. As noted above the instructions and output comparison files were prepared in Arc GIS Pro 3.1.1. Results will likely vary if using a different version of ArcGIS Pro.

5.1 Getting Started

5.1.1 Accessing the Training Data

Once the training data has been downloaded, note that the zipped folder contains the following:

Unzip the folders and save to a location of your choice The “Data” folder contains additional folders for each step of the process as follows:



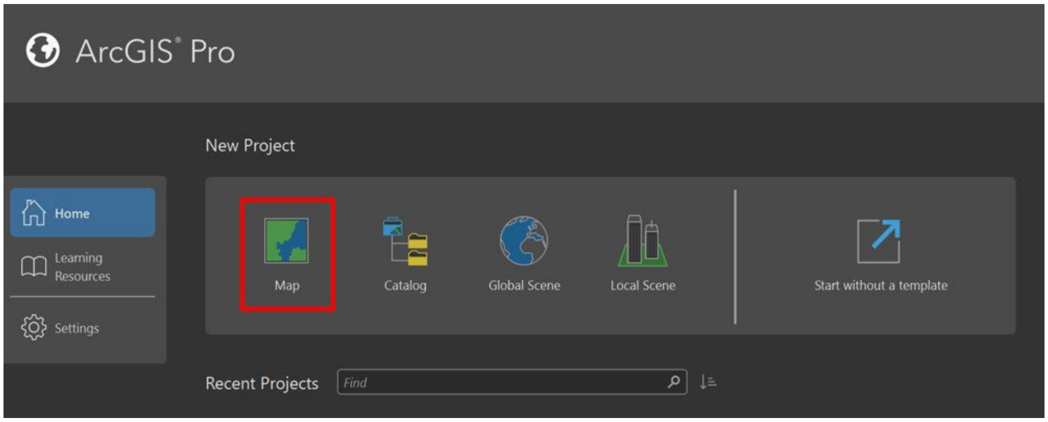
The Step_0 folder contains all the initial input data, including:

- Set_Classname.cal – used in data preparation (Section 5.2)
- Set_Classvalue.cal – used in data preparation (Section 5.2)
- Training_Data.shp – used in data preparation (Section 5.2) and classification (Section 5.4)
- Multispectral_Imagery.tif – used in segmentation (Section 5.3) and classification (Section 5.4)
- Reference_Data.tif – used in accuracy assessment (Section 5.6)

The folders also contain example outputs for each step and can be used for comparison to your results. Example files have the same names as the output files generated but are prefixed with “EX_” (e.g., “EX_Multispectral_Imagery_Seg.tif”) to prevent confusion between comparison files and generated output files.

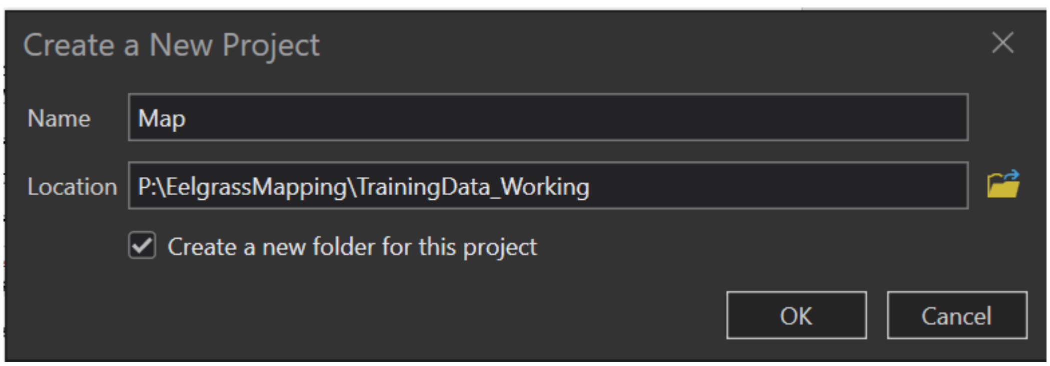
5.1.2 Creating the ArcGIS Pro Project

1. Open the ArcGIS Pro application, under New Project **Click-on** the "Map" icon to Create a New Project.

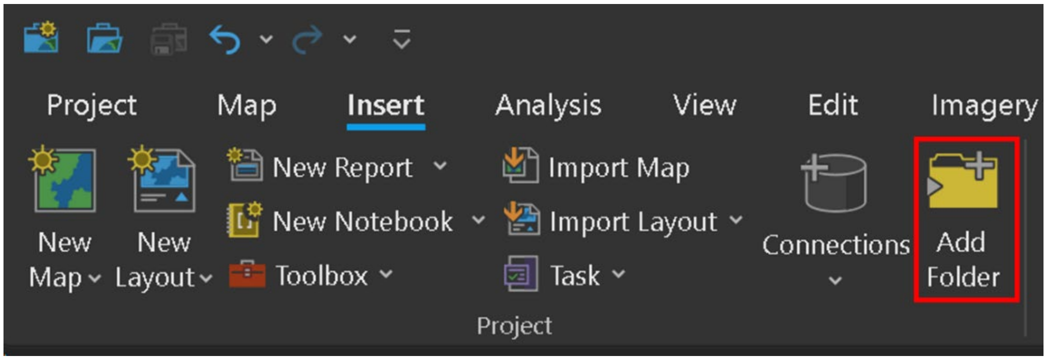


2. In the Create a New Project pop-up box, **Name** the project and find the location of your choice to save it, **Click** "OK". Make sure "Create a new folder for this project" is checked.

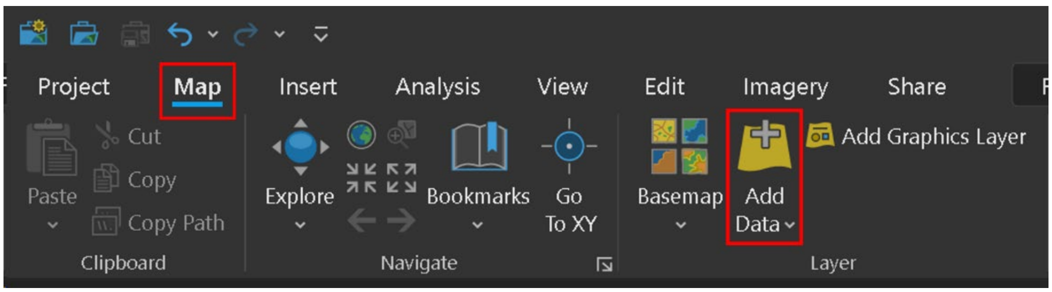
Note: The saved project file will have a .aprx extension and ArcGIS Pro will automatically create a project folder structure that includes a file geodatabase and ancillary files. Files created throughout this process will be saved to this default geodatabase and project folder. To learn more about how ArcGIS Pro organizes projects, go to the [Projects in ArcGIS Pro](#) website.



3. Create a folder connection to the unzipped training exercise data folder: In the top ribbon, **Go to** the "Insert" tab and **Click**-on "Add Folder". Find the unzipped training exercise folder and **Click** "OK".



4. Add data to the project: In the top ribbon, **Go to** the "Map" tab and **Click-on** "Add Data". **Find** the "Multispectral_Imagery.tif" file located in the Step_0_Input_Data folder, **Click** "OK". The image should appear in the Contents pane on the left.



5. ArcGIS Pro will zoom to the layer which is the study area (Beebe Cove).

6. Save your work: **Click-on** the "Save" button or **Go to** the "Project" tab on the top ribbon and **Click** "Save".

Note: Remember to save the project frequently to avoid losing your work due to software issues.

5.2 Preparing the Training Samples from Ground Truth Data (20-60 minutes)

As mentioned above, four of the five classification methods (all except ISO Cluster) require training samples to evaluate the imagery and classify accordingly (see Table 1). Training samples may be based on ground-truthing field data or created by drawing polygons using analyst knowledge of the imagery. For Beebe Cove, field data was collected within the same timeframe as the drone imagery with the purpose of supporting this classification effort. Figure 5.2a shows the sampling locations and attribute table for the ground-truthing samples collected at Beebe Cove. This field data was prepared for use as training samples for Beebe Cove.

If field data is not available, training samples can be created directly from the drone imagery. Steps to complete this process are included as Appendix A.

To use ground-truthing field data, samples must be assigned to representative classes (ex: eelgrass, bare sand, rock, other). In this case, the "Average Eelgrass (% cover)" attribute was used to assign samples to two classes (Eelgrass and Other). Any sample with an Average Eelgrass (% cover) value greater than or equal to 50 was assigned to the Eelgrass class, and any sample with a value less than 50 was assigned to the Other class. You will complete this step in Section 5.2.1 Formatting Samples for Classification.

Training_Data

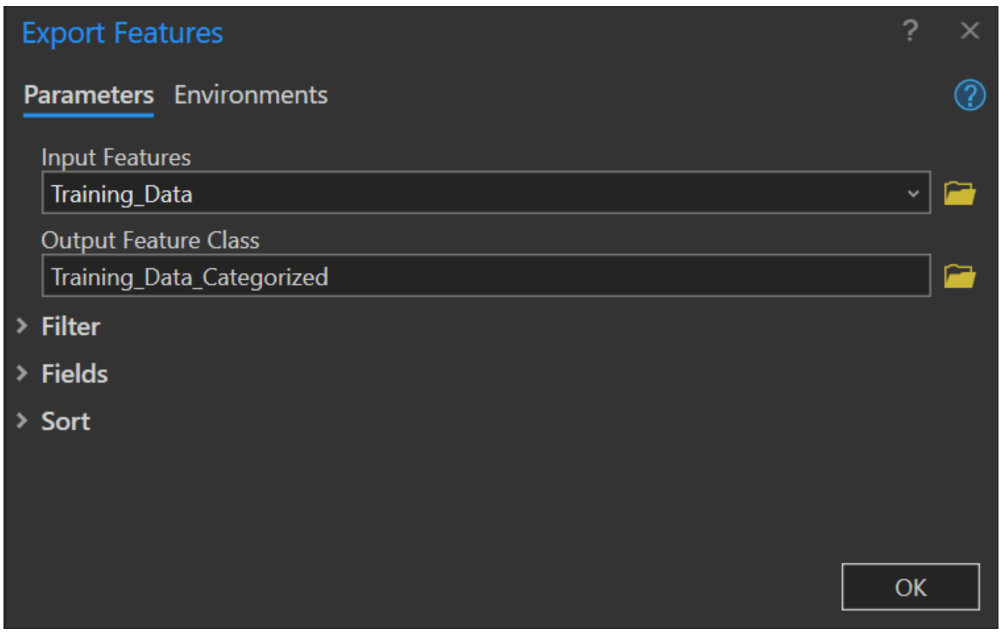
Field: Add Calculate Selection: Select By Attributes Zoom To Switch Clear Delete Copy

	FID	Shape *	Stn_ID	Latitude_	Longitude_	Average_Ee	Average_Ru	Average_Ma	Average_Ba	Epiphytes_	BUFF_DIST	ORIG_FID	Shape_Leng	Shape_Area
1	0	Polygon	1	41.3299	-71.9883	94	3.14	0.43	2.43	1.29	30	1	188.495559	2827.433388
2	1	Polygon	2	41.33012	-71.98718	97.83	1.83	0	0	2.83	30	2	188.495559	2827.433388
3	2	Polygon	3	41.33043	-71.98618	0	0	95.67	4.33	0	30	3	188.495559	2827.433388
4	3	Polygon	4	41.33115	-71.98539	0	0	71.57	22.43	0	30	4	188.495559	2827.433388
5	4	Polygon	5	41.33094	-71.98624	100	0	0	0	3	30	5	188.495559	2827.433388
6	5	Polygon	6	41.33072	-71.98697	100	0	0	0	3	30	6	188.495559	2827.433388
7	6	Polygon	7	41.33052	-71.98792	99.6	0.4	0	0	1.2	30	7	188.495559	2827.433388
8	7	Polygon	8	41.33022	-71.9887	86.8	1.2	0	12	1.6	30	8	188.495559	2827.433388
9	8	Polygon	9	41.33081	-71.98915	95.8	0	0	4.2	2	30	9	188.495559	2827.433388
10	9	Polygon	10	41.33073	-71.98943	5	3	30.8	61.2	1	30	10	188.495559	2827.433388
11	10	Polygon	11	41.33113	-71.98729	100	0	0	0	3	30	11	188.495559	2827.433388
12	11	Polygon	12	41.33135	-71.98627	0	0	67.8	32.2	0	30	12	188.495559	2827.433388
13	12	Polygon	13	41.3315	-71.98552	22	0	43	35	2	30	13	188.495559	2827.433388
14	13	Polygon	14	41.3321	-71.98544	15.8	0	30.2	54	1.6	30	14	188.495559	2827.433388
15	14	Polygon	15	41.33188	-71.9861	40	0	1.2	58.8	3	30	15	188.495559	2827.433388

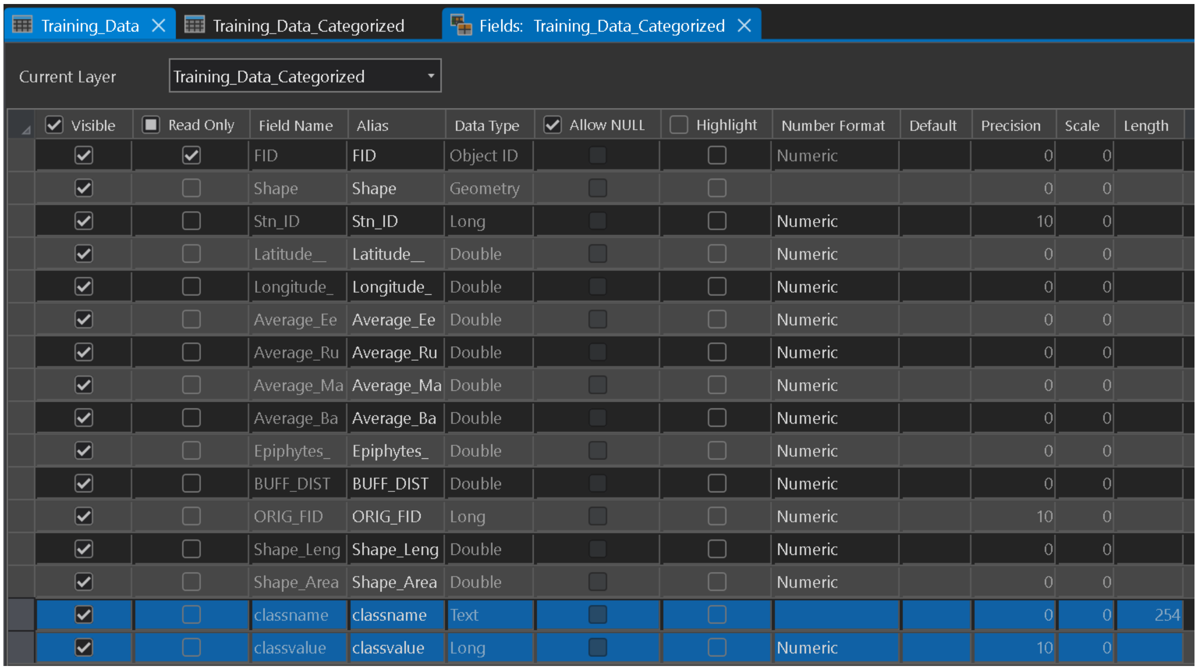
Figure 5.2a – Ground-Truthing sample locations and attribute table for field data collected at Beebe Cove.

5.2.1 Formatting Training Samples for Classification

1. **Add** Training_Data.shp from the Step_0_Input_Data file to the ArcGIS Pro Contents pane.
2. Create a copy of the training data: In the Contents pane, **Right-Click** on the shapefile (Training_Data), **Go to** "Data" and **Select** "Export Features".
3. **Name** the "Output Feature Class" Training_Data_Categorized and save to the default project map geodatabase. The new file will appear in the Contents pane.



4. Add the necessary fields required for classification: **Right-Click** on Training_Data_Categorized in the Contents pane and **Open** the Attribute Table.
5. Insert a column for classname: In the Attribute Table, **Click-on** "Add" next to "Field:" The "Fields Tab" will open. **Type** in "classname" in both the "Field Name" and "Alias" columns and **Select** "Text" from the "Data Type" drop down menu. **Close** the "*Fields: Training_Data_Categorized" tab. **Save** the changes and **Confirm** the new "classname" column in the Attribute Table.
6. Insert a column for classvalue: In the Attribute Table, **Click-on** "Add" next to "Field:" The "Fields Tab" will open. **Type** in "classvalue" in both the "Field Name" and "Alias" columns and **Select** "Long" from the "Data Type" drop down menu. **Close** the "*Fields: Training_Data_Categorized" tab. **Save** the changes and **Confirm** the new "classvalue" column in the Attribute Table.



Visible	Read Only	Field Name	Alias	Data Type	Allow NULL	Highlight	Number Format	Default	Precision	Scale	Length
✓	✓	FID	FID	Object ID			Numeric		0	0	
✓		Shape	Shape	Geometry					0	0	
✓		Stn_ID	Stn_ID	Long			Numeric		10	0	
✓		Latitude_	Latitude_	Double			Numeric		0	0	
✓		Longitude_	Longitude_	Double			Numeric		0	0	
✓		Average_Ee	Average_Ee	Double			Numeric		0	0	
✓		Average_Ru	Average_Ru	Double			Numeric		0	0	
✓		Average_Ma	Average_Ma	Double			Numeric		0	0	
✓		Average_Ba	Average_Ba	Double			Numeric		0	0	
✓		Epiphytes_	Epiphytes_	Double			Numeric		0	0	
✓		BUFF_DIST	BUFF_DIST	Double			Numeric		0	0	
✓		ORIG_FID	ORIG_FID	Long			Numeric		10	0	
✓		Shape_Leng	Shape_Leng	Double			Numeric		0	0	
✓		Shape_Area	Shape_Area	Double			Numeric		0	0	
✓		classname	classname	Text					0	0	254
✓		classvalue	classvalue	Long			Numeric		10	0	

7. Enter the information related to the class name and class value for all the training samples.

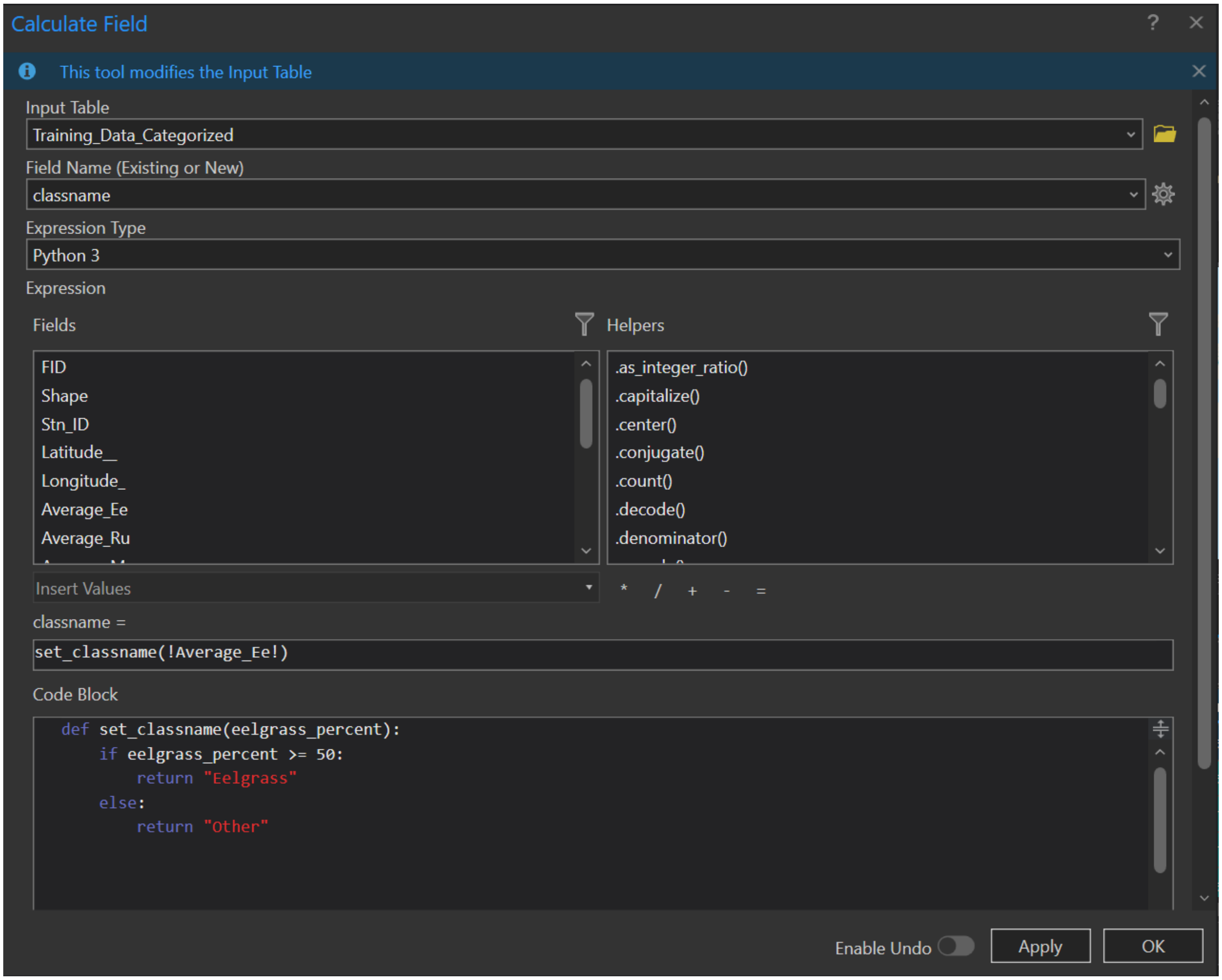
Note: you can enter information for each attribute manually, but for this training exercise, you will use pre-saved expressions, one for classname and the other for classvalue. The advantage of the expressions is that they populate fields automatically, saving time and eliminating errors that may occur through manual entry.

a. In the "Training_Data_Categorized Attribute Table, **Right-click** on the "classname" field and **Select** "Calculate Field". This opens the "Calculate Field" window. See the below screenshot for input details.

i. Below the "Code Block", **Click** on the folder icon and **Select** the "Set_Classname.cal" file (from Step_0_Input_Data), **Click** "OK".

Note that this file contains an expression and subsequent code to populate the class name based on the eelgrass percentage discussed above.

ii. **Click** "Apply" to populate the field, then "OK" to exit the window.

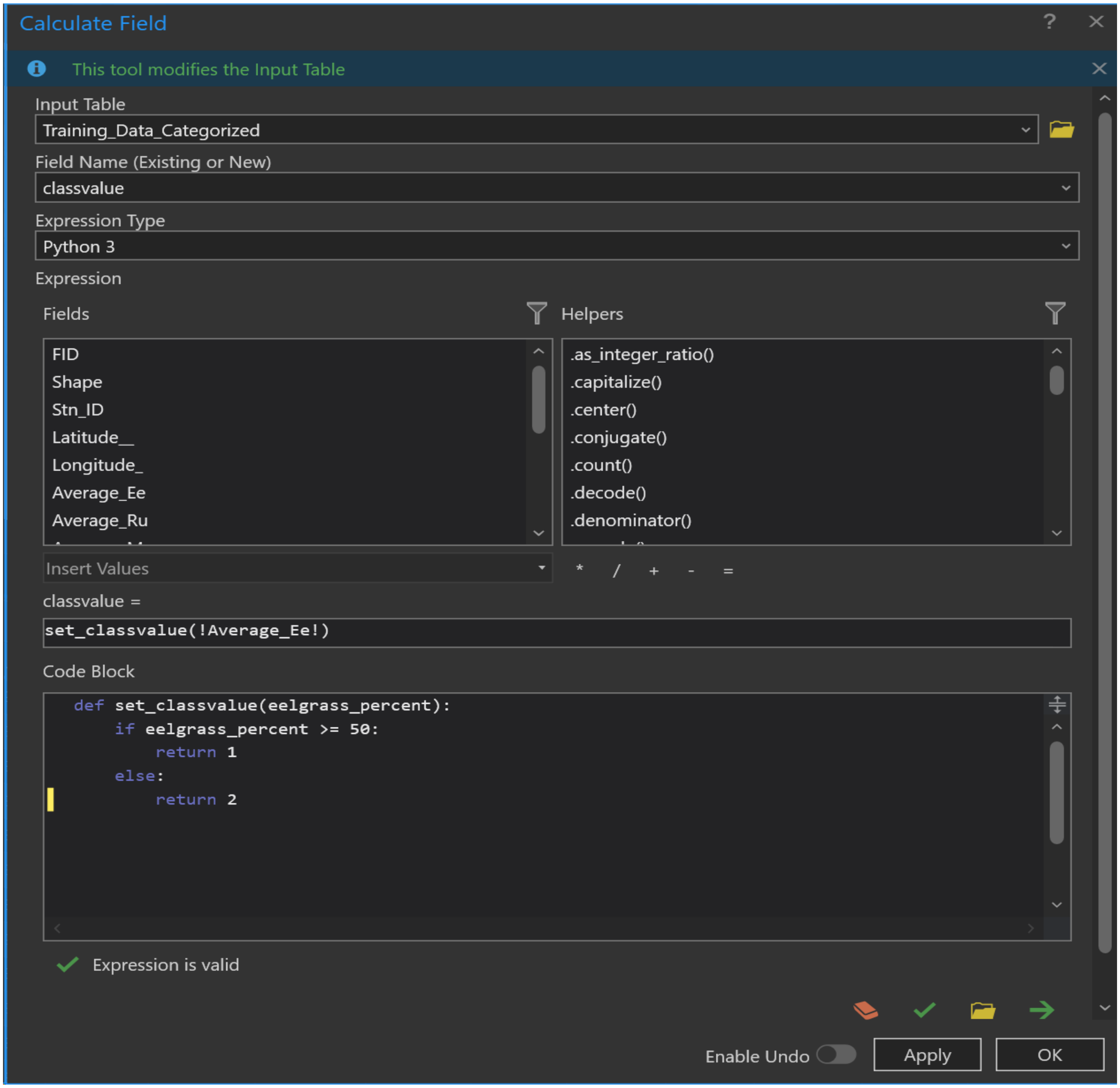


b. Next, **Right-click** on the “classvalue” field and **Select** “Calculate Field”. This opens the “Calculate Field” window. See the below screenshot for input details.

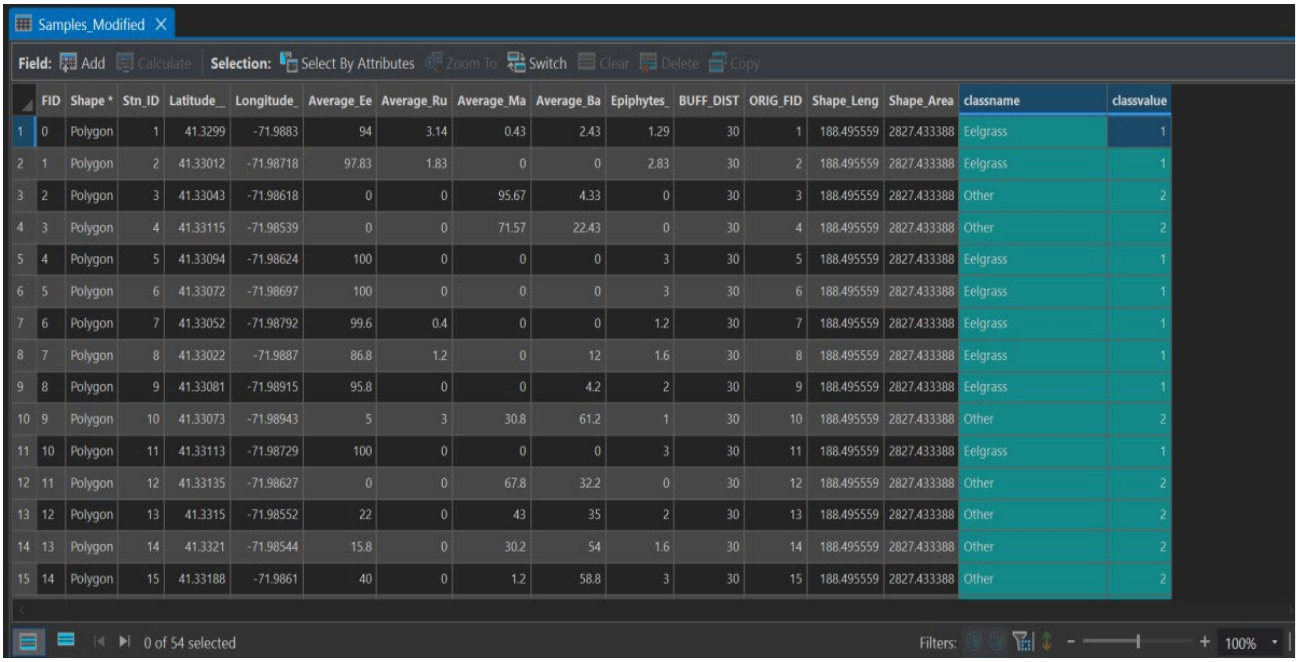
i. Below the “Code Block”, **Click** on the folder icon and **Select** the “Set_Classvalue.cal” file (from Step_0_Input_Data), **Click** “OK”.

Note that this file contains an expression and subsequent code to populate the class value based on the Eelgrass percentage discussed above.

ii. **Click** “Apply” to populate the field, then “OK” to exit the window.



Once the classname and classvalue fields are populated with the correct information, the shapefile can be used in the classification. The following screenshots show dataset examples with these attributes populated, with two classes: Eelgrass (1) and Other (2).



	FID	Shape *	Stn_ID	Latitude_	Longitude_	Average_Ee	Average_Ru	Average_Ma	Average_Ba	Epiphytes_	BUFF_DIST	ORIG_FID	Shape_Leng	Shape_Area	classname	classvalue
1	0	Polygon	1	41.3299	-71.9883	94	3.14	0.43	2.43	1.29	30	1	188.495559	2827.433388	Eelgrass	1
2	1	Polygon	2	41.33012	-71.98718	97.83	1.83	0	0	2.83	30	2	188.495559	2827.433388	Eelgrass	1
3	2	Polygon	3	41.33043	-71.98618	0	0	95.67	4.33	0	30	3	188.495559	2827.433388	Other	2
4	3	Polygon	4	41.33115	-71.98539	0	0	71.57	22.43	0	30	4	188.495559	2827.433388	Other	2
5	4	Polygon	5	41.33094	-71.98624	100	0	0	0	3	30	5	188.495559	2827.433388	Eelgrass	1
6	5	Polygon	6	41.33072	-71.98697	100	0	0	0	3	30	6	188.495559	2827.433388	Eelgrass	1
7	6	Polygon	7	41.33052	-71.98792	99.6	0.4	0	0	1.2	30	7	188.495559	2827.433388	Eelgrass	1
8	7	Polygon	8	41.33022	-71.9887	86.8	1.2	0	12	1.6	30	8	188.495559	2827.433388	Eelgrass	1
9	8	Polygon	9	41.33081	-71.98915	95.8	0	0	4.2	2	30	9	188.495559	2827.433388	Eelgrass	1
10	9	Polygon	10	41.33073	-71.98943	5	3	30.8	61.2	1	30	10	188.495559	2827.433388	Other	2
11	10	Polygon	11	41.33113	-71.98729	100	0	0	0	3	30	11	188.495559	2827.433388	Eelgrass	1
12	11	Polygon	12	41.33135	-71.98627	0	0	67.8	32.2	0	30	12	188.495559	2827.433388	Other	2
13	12	Polygon	13	41.3315	-71.98552	22	0	43	35	2	30	13	188.495559	2827.433388	Other	2
14	13	Polygon	14	41.3321	-71.98544	15.8	0	30.2	54	1.6	30	14	188.495559	2827.433388	Other	2
15	14	Polygon	15	41.33188	-71.9861	40	0	1.2	58.8	3	30	15	188.495559	2827.433388	Other	2

classname	classvalue
Eelgrass	1
Eelgrass	1
Other	2
Other	2
Eelgrass	1
Eelgrass	1
Eelgrass	1
Eelgrass	1
Eelgrass	1
Other	2
Eelgrass	1
Other	2
Other	2
Other	2
Other	2

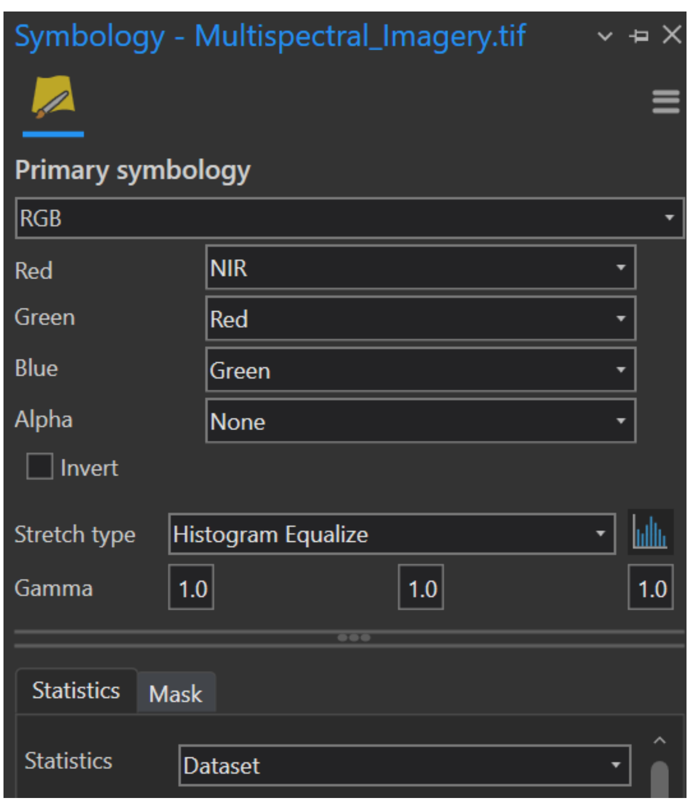
Close the attribute table, **Save** the project, and **Go to** Section 5.3.

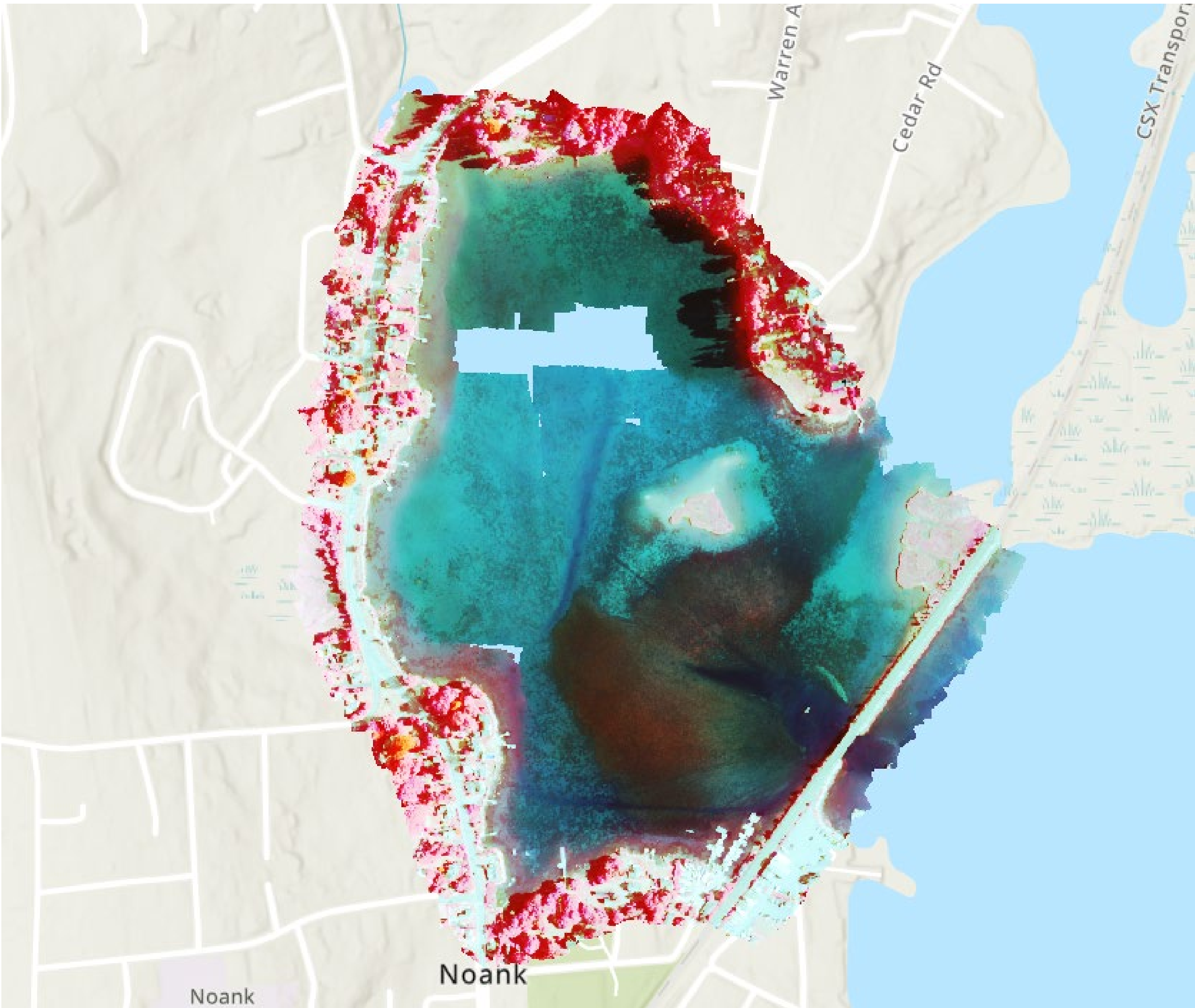
5.3 Image Segmentation (45 - 150 Minutes)

For object-based classification, a segmented image must be created prior to executing the classification methods. Applying segmentation to the Beebe Cove orthomoasic creates image segments/objects (groups of pixels) that will be categorized into Eelgrass and Other. This is compared to pixel-based classifications, where pixels are individually categorized as opposed to being grouped into segments/objects. As mentioned above, this training exercise covers object-based classification methods.

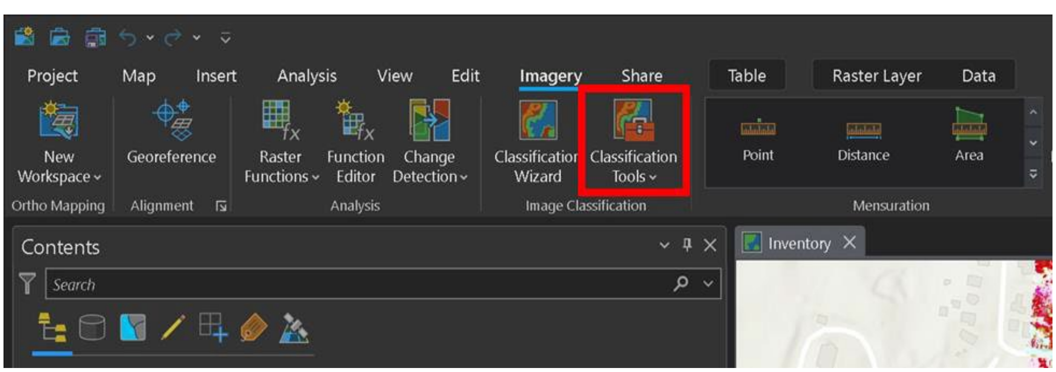
1. In the Contents pane, **Right-click** on Multispectral_Imagery.tif and **Select** "Zoom to Layer".
2. Adjust the image symbology so the features for classification (i.e. Eelgrass and Other) are visually clear:
    a. In the Contents pane, **Right-click** on the layer and Select "Symbology". The Symbology pane will open.
    b. **Set** the "Primary symbology" to RGB, "Red" to NIR, "Green" to Red, and "Blue" to Green.
    c. In the "Stretch type", **Select** Histogram Equalize.
    d. In the "Statistics" tab**, Select** Dataset.

Note: These settings provide the best ability to see and delineate the eelgrass from the background. See the following screenshots for symbology settings and final visualization.

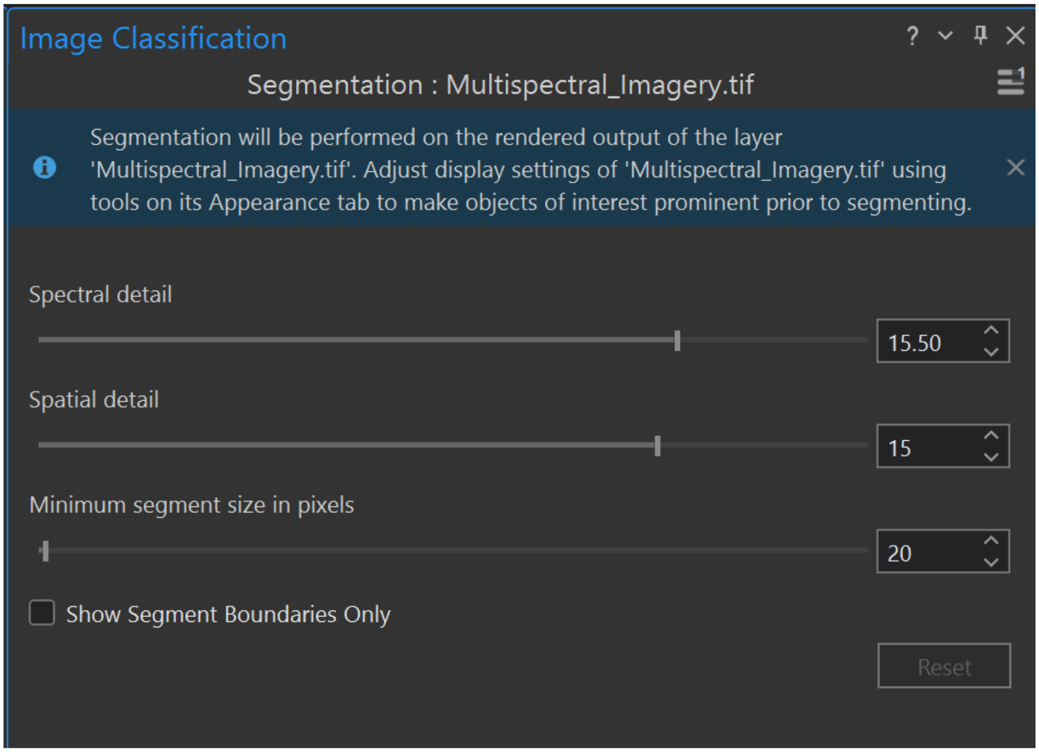




3. Create the Segmented Image (geotiff): In the main Contents pane, **Select** the symbolized image (Multispectral_Imagery). In the ArcGIS Pro ribbon, **Click** the "Imagery tab", **Select** "Classification Tools", **Select** "Segmentation". The Image Classification pane will open.

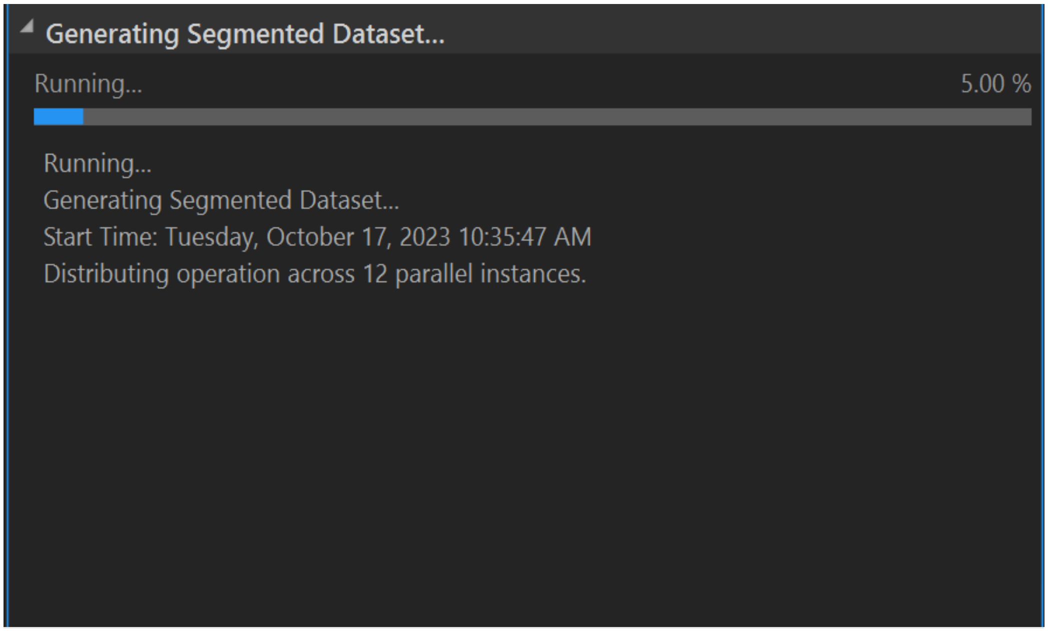


4. In the Image Classification pane, **Observe** the Segmentation Settings. For Beebe Cove, the default values were used and are shown in the below screenshot. More information about segmentation parameters can be found on ESRI's website.



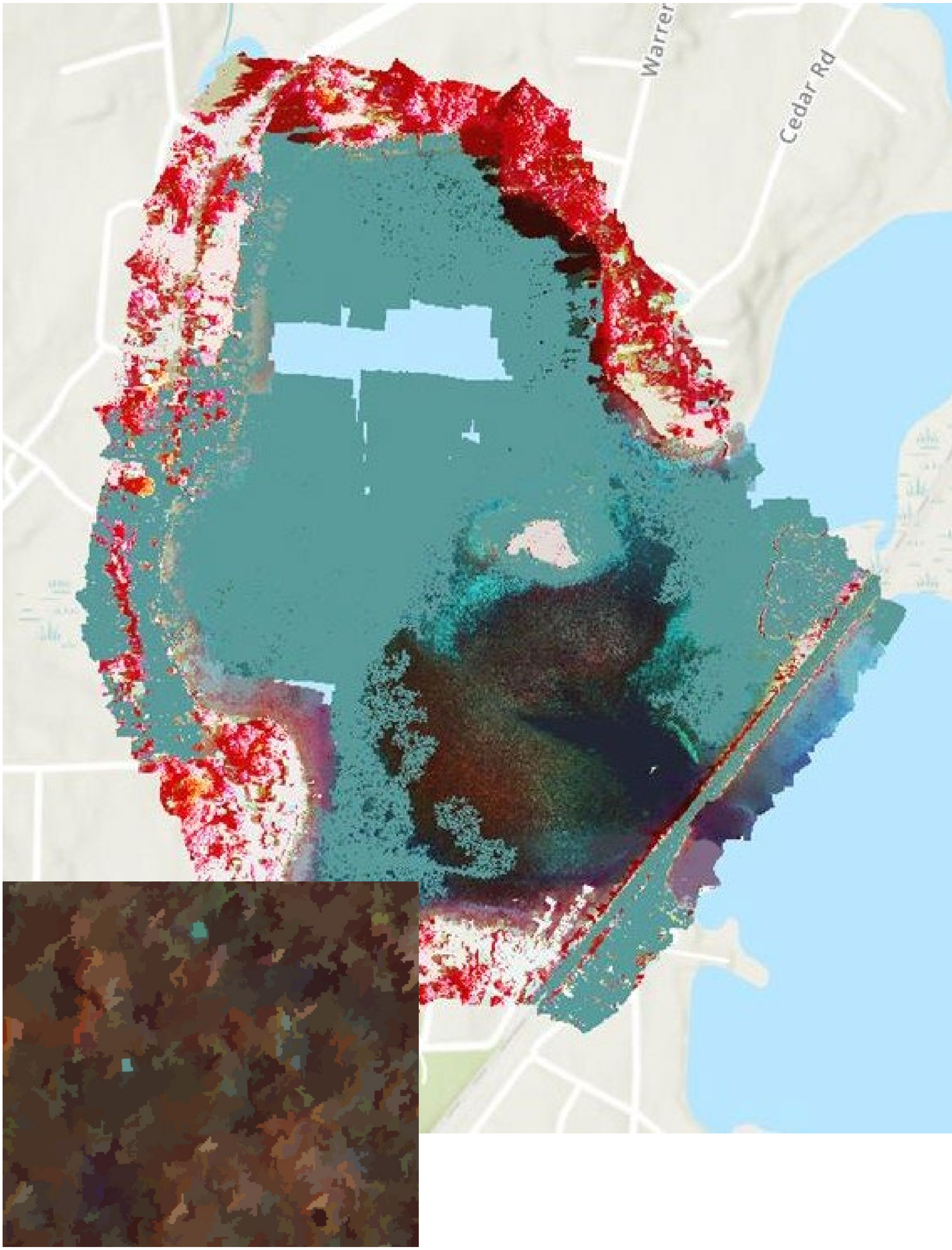
5. **Name** the "Output Dataset" Multispectral_Imagery_Segmented and **Save** to the default project geodatabase.

6. **Check** that all fields are correctly populated and **Click** "Run". A progress bar will appear which can be used to track the completion of this process.



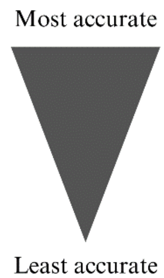
Once segmentation is complete, the segmented image (Multispectral_Imagery_Segmented) will be auto added to the ArcGIS map. This segmented image will have a similar visualization to the input image. A screenshot of the segmented image is shown below. This full site image is at scale of 1:10,000. An additional image at the scale of 1:50 scale is included to show the individual segments.

If you would like to view an example of the results in ArcGIS, you can add the output of this tool to your map by going to the Map tab, Select Add Data, and Toggle to the folder: Step_2_Image_Segmentation. Select the file called EX_Multispectral_Imagery_Segmented.tif.


Cedar Rd

5.4 Image Classification (30-120 minutes)

In ArcGIS Pro Classification Tools, five classification methods are available. When run with the Beebe Cove imagery, the perceived level of accuracy for the object-based classification methods was as follows:



Maximum Likelihood

Random Trees

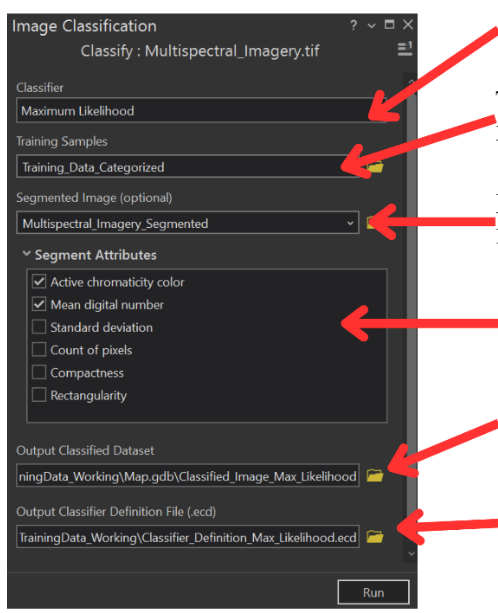
Support Vector Machine

K-Nearest Neighbor

ISO Cluster

Though five classification methods are available, this training exercise only provides steps on executing the Maximum Likelihood method. Additional information on classification tools and parameters can be accessed on ESRI's website: Classify

Required fields common to the object-based classification methods include training samples, segmented image, and segment attributes. Figure 5.4a describes the input files, settings, and output files.



Classification method

Training Samples input, created in **Section 5.2**.
For all methods except ISO Cluster, the "Training Samples" field must be used.

Image segments/objects input, created in **Section 5.3**.
In order to do an object-based classification, you need to include a segmented image here.

Settings to adjust the classification
For this exercise, default settings were used.

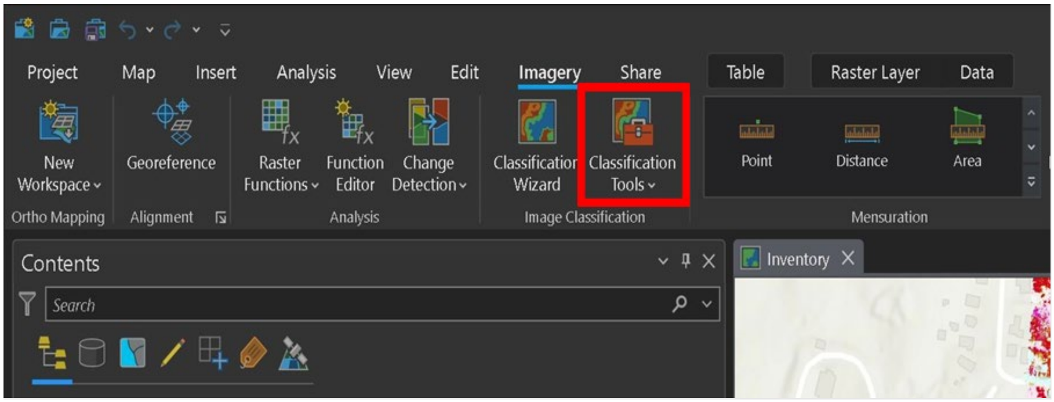
Name of classified dataset output
Created during the classification process. The user is only responsible for naming this.

Name of classifier definition file output
Created during the classification process. The user is only responsible for naming this.

Figure 5.4a – Image Classification Screenshot with descriptions of the input files, settings, and output files.

5.4.1 Executing the Maximum Likelihood Classification Method

1. In the main Contents pane, **Select** the original image (Multispectral_Imagery.tif), In the ArcGIS Pro ribbon, **Click** the "Imagery tab", **Click** "Classification Tools". From the drop-down menu, **Select** "Classify". This will launch the Image Classification pane.



2. Populate the Image Classification pane as follows:

    a. Under "Classifier" in the Image Classification pane, **Select** Maximum Likelihood from the drop-down menu.

    b. For "Training Samples", **Select** the Training_Data_Categorized file from the drop-down menu.

    c. For "Segmented Image", **Select** the Multispectral_Imagery_Segmented file from the drop-down menu.

    d. In the "Segment Attributes" box, **Check** the boxes for "Active chromaticity color" and "Mean digital number".

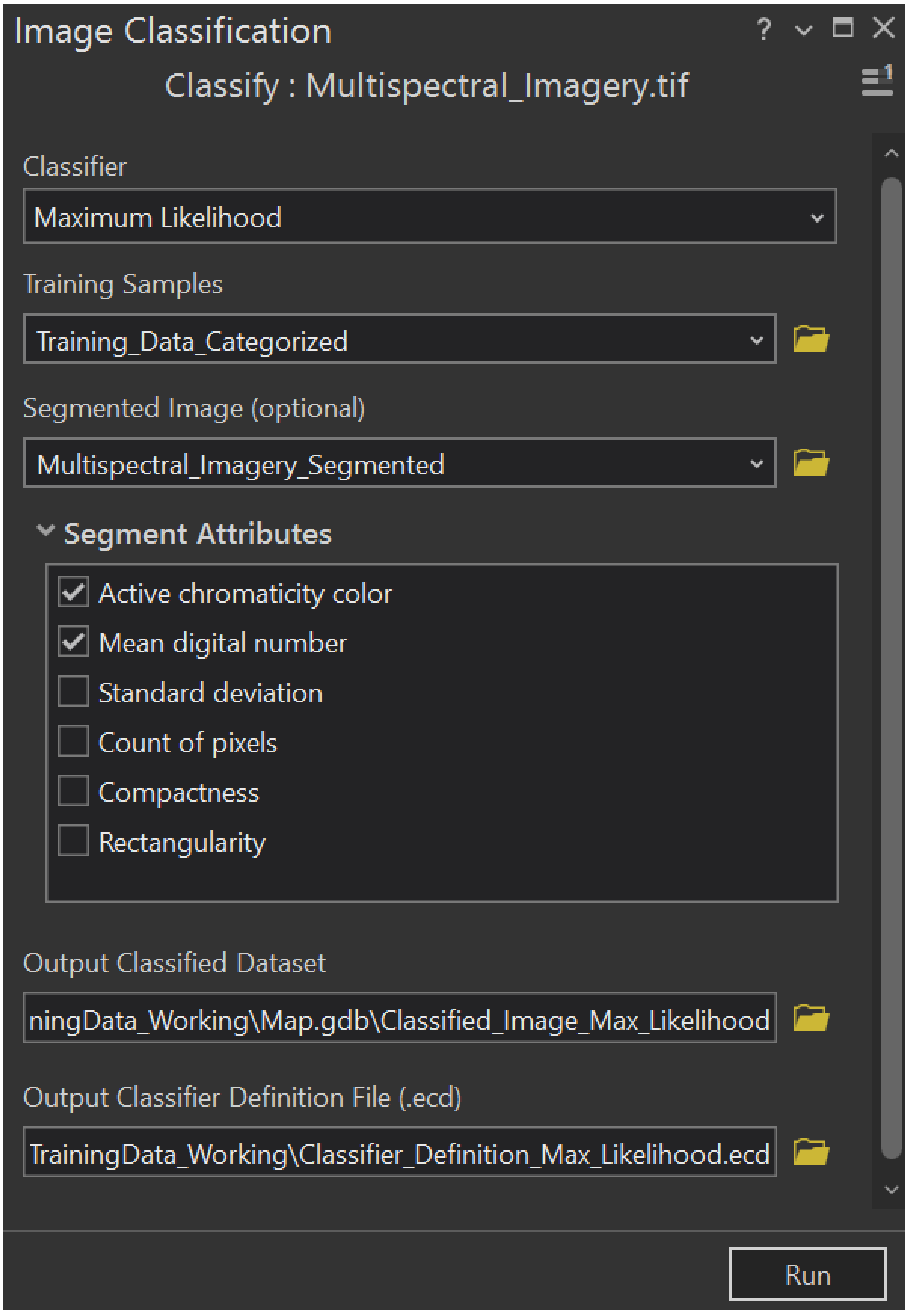
    Note: standard deviation could also be selected, however, it did not cause a substantial difference for this exercise.

    e. **Name** the "Output Classified Dataset" Classified_Image_Max_Likelihood and **Save** to the default project geodatabase.

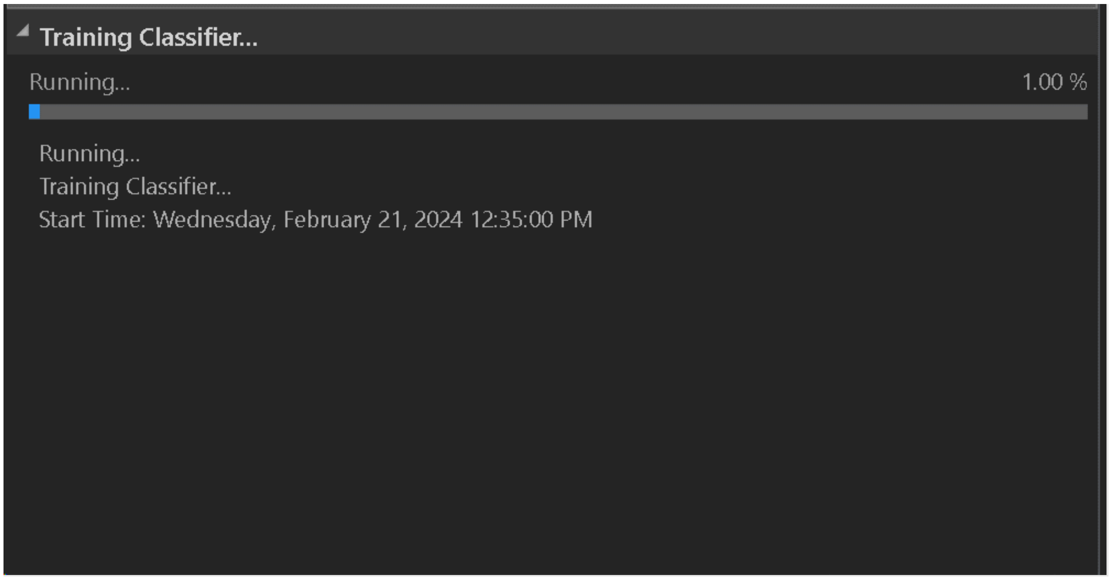
    f. **Name** the "Output Classifier Definition File" Classifier_Definition_Max_Likelihood and **Save** to the default project folder.

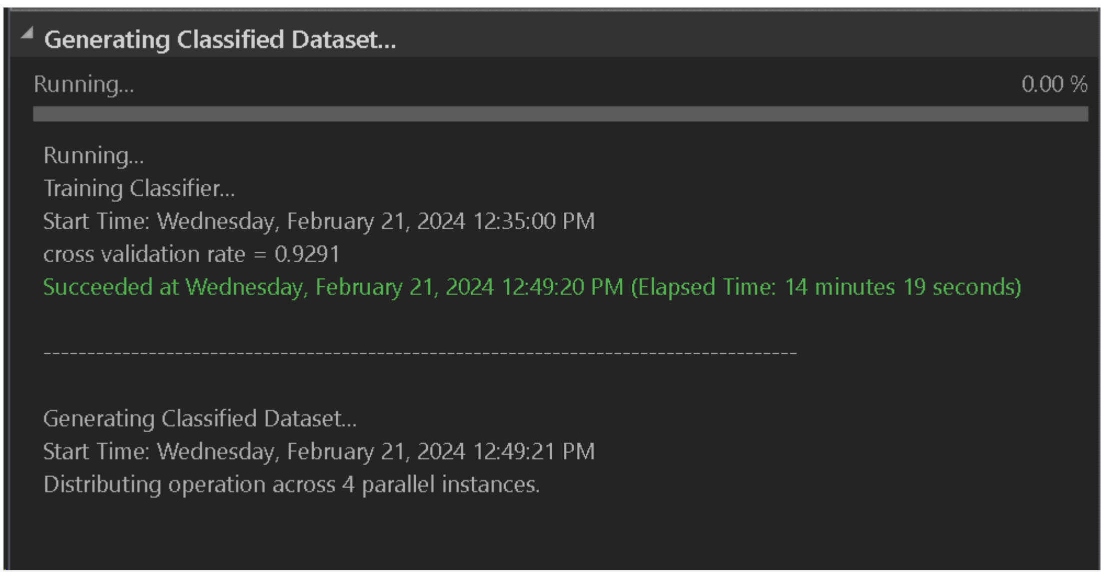
3. **Check** that all fields are correctly populated using the following screenshot.

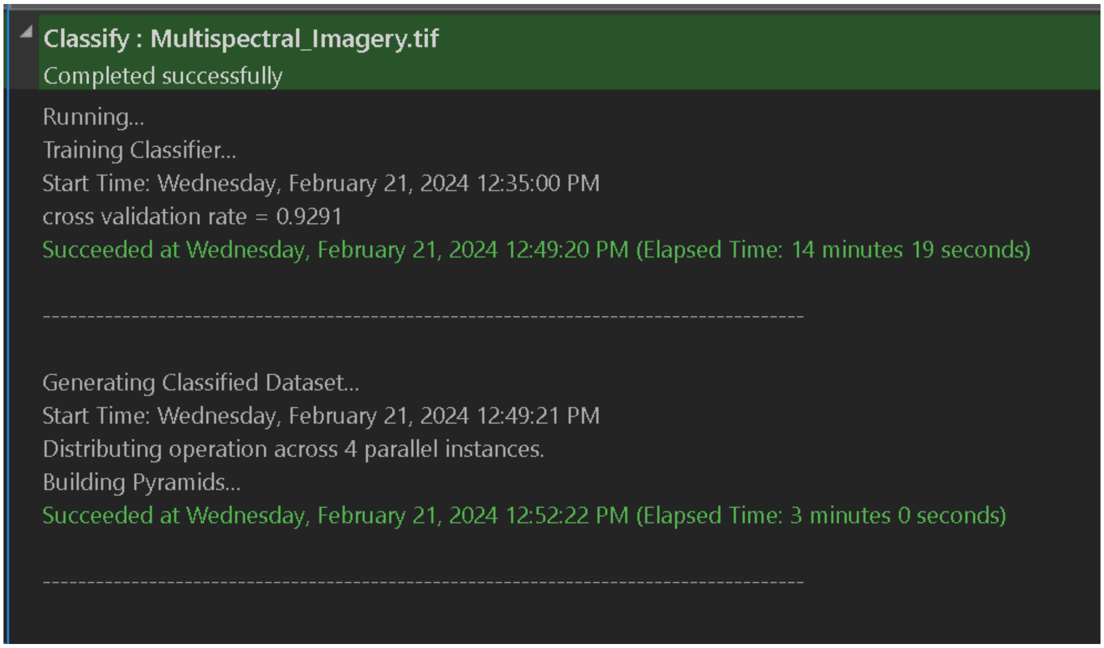
Note: Files included in the main Contents pane will be available in the Image Classification drop-down menu for the Training Samples and the Segmented Image fields. If they are not available, they can be added by clicking on the folder icon and toggling to their saved location.



4. Once you've confirmed that all fields are correctly populated, **Click** "Run". A progress bar will appear which can be used to track the completion of this process. The "Running" window also shows current and completed steps, along with ancillary information associated with a particular step (e.g., number of parallel instances used during processing, seen during the "Generating Classified Dataset" step). Text will appear once processing has completed, and the window will show the completion time.

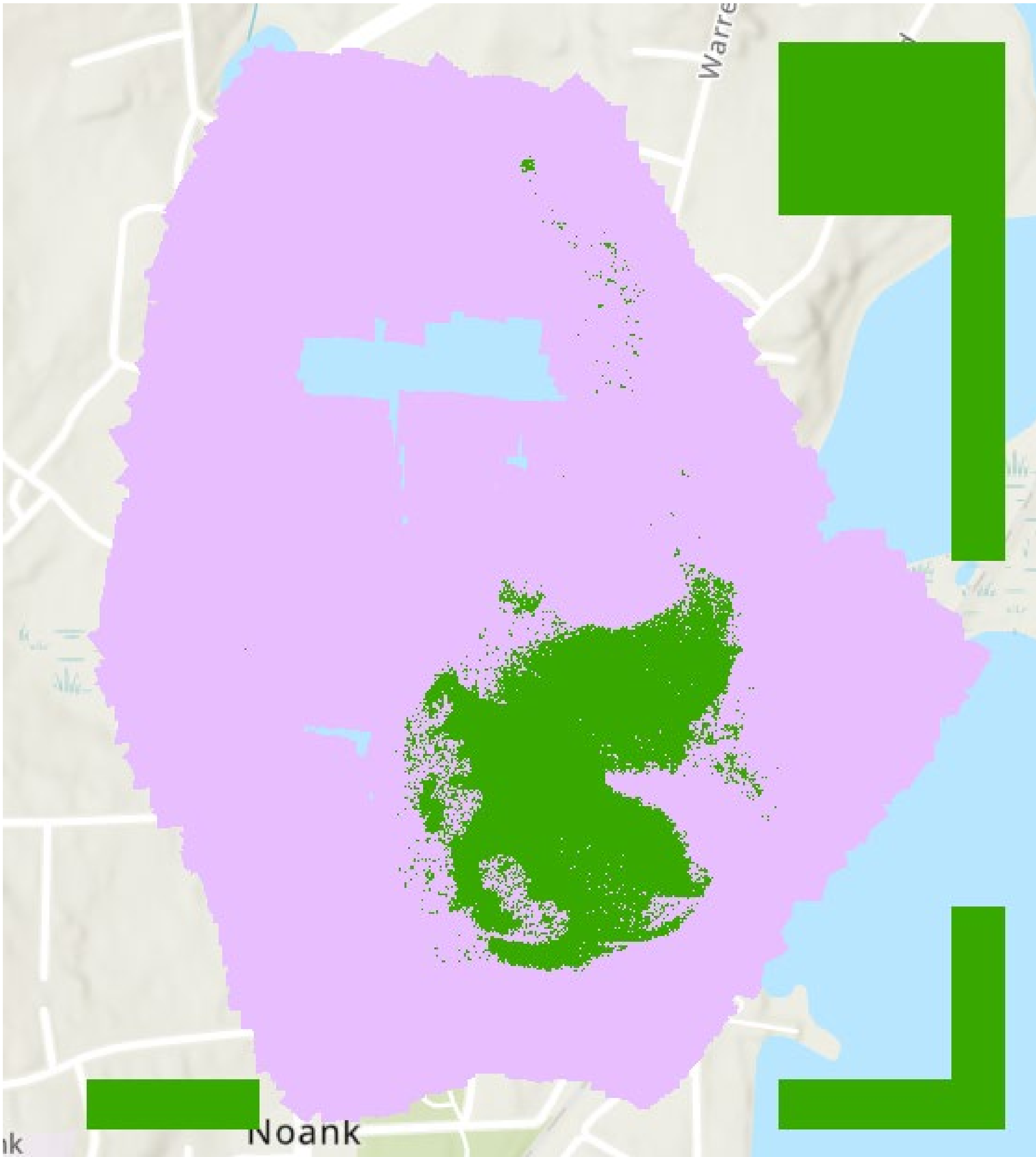

Training Classifier...
Running...
1.00 %
Running...
Training Classifier...
Start Time: Wednesday, February 21, 2024 12:35:00 PM


Generating Classified Dataset...
Running...
0.00 %
Running...
Training Classifier...
Start Time: Wednesday, February 21, 2024 12:35:00 PM
cross validation rate = 0.9291
Succeeded at Wednesday, February 21, 2024 12:49:20 PM (Elapsed Time: 14 minutes 19 seconds)
Generating Classified Dataset...
Start Time: Wednesday, February 21, 2024 12:49:21 PM
Distributing operation across 4 parallel instances.



Once complete, the Output Classified Dataset (Classified_Image_Max_Likelihood) will auto load into the ArcGIS map. The example screenshot below shows the result of the Maximum Likelihood classification and can be used for comparison to your output classified dataset. “Eelgrass” features are symbolized in green, and the “Other” features are symbolized in purple. If the symbology is not as shown, adjust it as follows:

1. In the Contents pane, **Right-click** on the Classified_Image_Max_Likelihood file and Select “Symbology”. The Symbology pane will open.
2. In the Primary symbology drop down, **Select** “Unique Values”.
3. In the “Field 1” drop down, **Select** Class_name.
4. In the Values tab, **Select** the color block next to “Eelgrass”. **Click on** “Color Properties...” and in the HEX # box **Enter** 38A800. **Click** “OK”.
5. In the Values tab, **Select** the color block next to “Other”. **Click on** “Color Properties...” and in the HEX # box **Enter** E8BEFF. **Click** “OK”.


Noank

5.5 Accuracy Assessment (10-30 minutes)

5.5.1 Introduction

A remote sensing accuracy assessment is a useful metric for a quantitative estimate of the classification accuracy, via a confusion/error matrix. This matrix shows user's accuracy (measure of commission errors), producer's accuracy (measure of omission errors), and overall accuracy, with a table comparing the classified data to reference data. Esri has put together a video demonstration, which includes the following steps:

- Create accuracy assessment points
- Update accuracy assessment points
- Compute confusion matrix

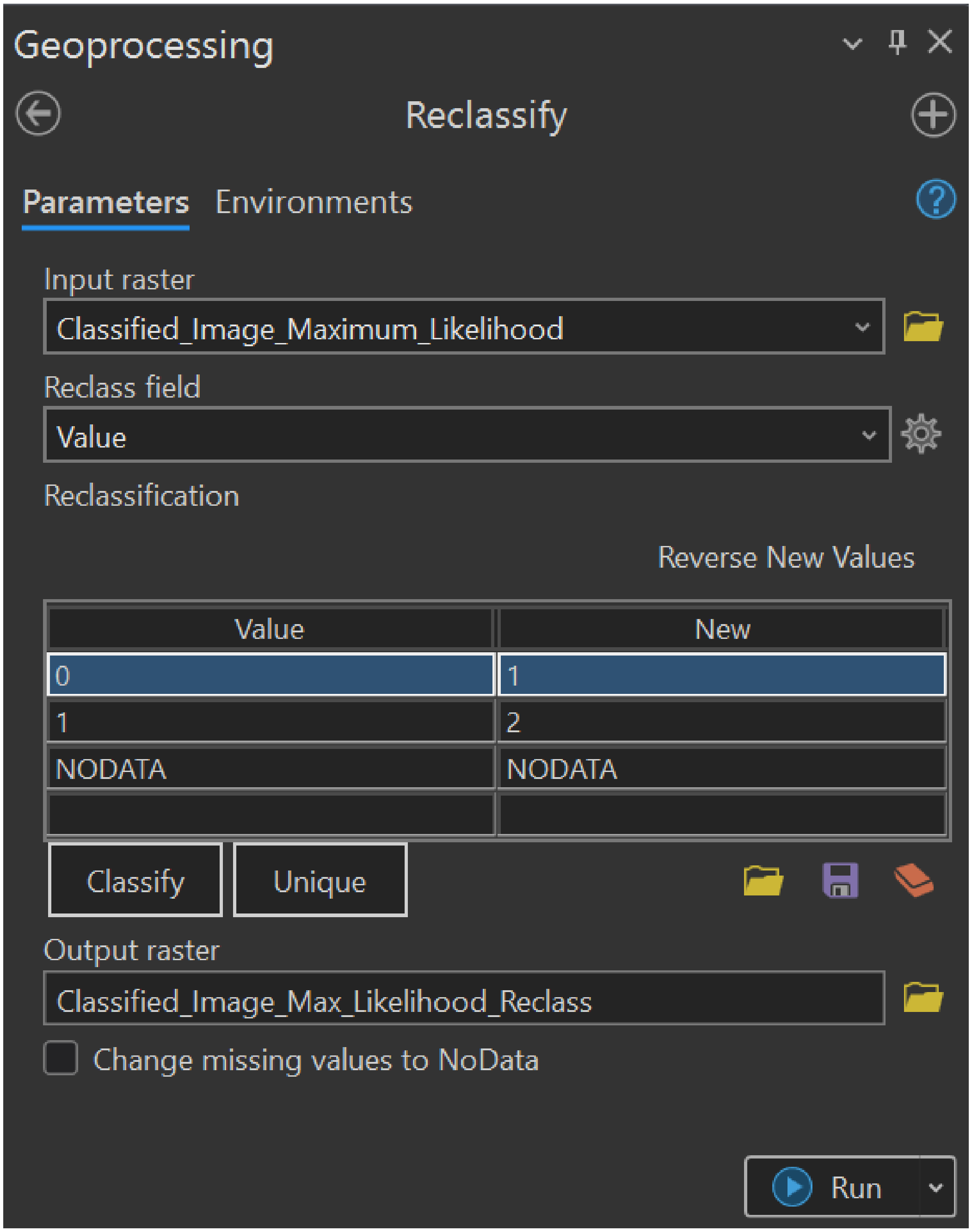
Additionally, an overview of the process can be found here.

In general, the accuracy assessment points should be different than the training samples. It is best to have different sets of samples used specifically for training and validation, where the validation samples are used for the accuracy assessment points. This prevents the use of the training samples as the accuracy assessment points, which can lead to higher (and potentially misleading) accuracy numbers. If enough samples are available for both training and validation, it's recommended to split the samples 80% for training and 20% for validation.

5.5.2 Prepare Raster for Accuracy Assessment

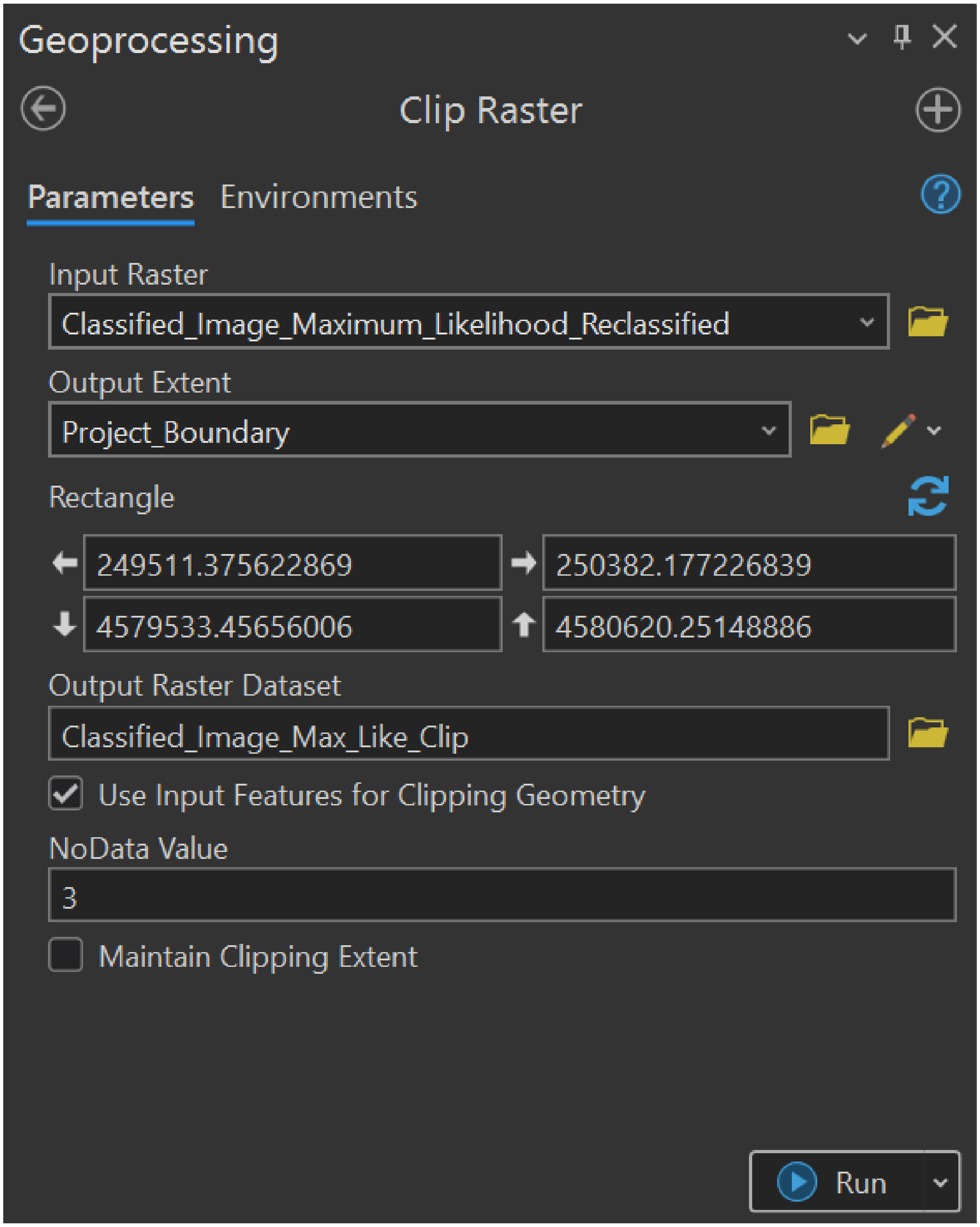
For this exercise, you will use Reference_Data.tif, which is a rasterized version of the manually digitized eelgrass features for Beebe Cove, as the reference data, which was previously created. In order for the accuracy assessment to succeed, the classified raster and the reference raster must be in the same format. This requires an adjustment to the values of the classified raster.

1. Click on the Geoprocessing Tools box in the Analysis Ribbon. The Geoprocessing pane will open and search for the reclassify tool. Open the **Reclassify** (Spatial Analyst Tools) geoprocessing tool.
    a. For "Input raster", **Select** the Classified_Image_Max_Likelihood file from the drop-down menu.
    b. For "Reclass field", **Select** Value from the drop-down menu.
    c. **Click-on** the "Unique" button to populate the "Value" column.
    d. Set the "New" column to "1", "2", and "NODATA" as shown in the following screenshot.
    e. **Name** it Classified_Image_Max_Likelihood_Reclass and **Save** the "Output raster" to the default project geodatabase.
    f. **Click** the Run button.


Geoprocessing
Reclassify
Parameters
Environments
Input raster
Classified_Image_Maximum_Likelihood
Reclass field
Value
Reclassification
Reverse New Values
Value
New
0
1
1
2
NODATA
NODATA
Classify
Unique
Output raster
Classified_Image_Max_Likelihood_Reclass
Change missing values to NoData
Run

In addition to the reclassification, and in order to prevent NoData values from being classified into Eelgrass or Other, clip the reclassified Maximum Likelihood raster to the project boundary prior to conducting the accuracy assessment steps.

1. In the Geoprocessing pane, **Search** for the Clip Raster tool and **Select** the Clip Raster (Data Management Tools) option.
    a. For "Input Raster", **Select** the Classified_Image_Max_Likelihood_Reclass file from the drop-down menu.
    b. For the "Output Extent", **Select** the Project_Boundary file from the drop-down menu. This will automatically populate the "Rectangle" parameters.
    c. **Name** it Classified_Image_Max_Like_Reclass_Clip and **Save** the "Output raster" to the default project geodatabase.
    d. **Click** on the "Use Input Features for Clipping Geometry" box (this is unchecked by default).
    e. The "NoData Value" parameter should be listed as 3 or adjust it to 3.
    f. **Confirm** your settings with the Clip Raster screenshot below and **Click** the Run button.

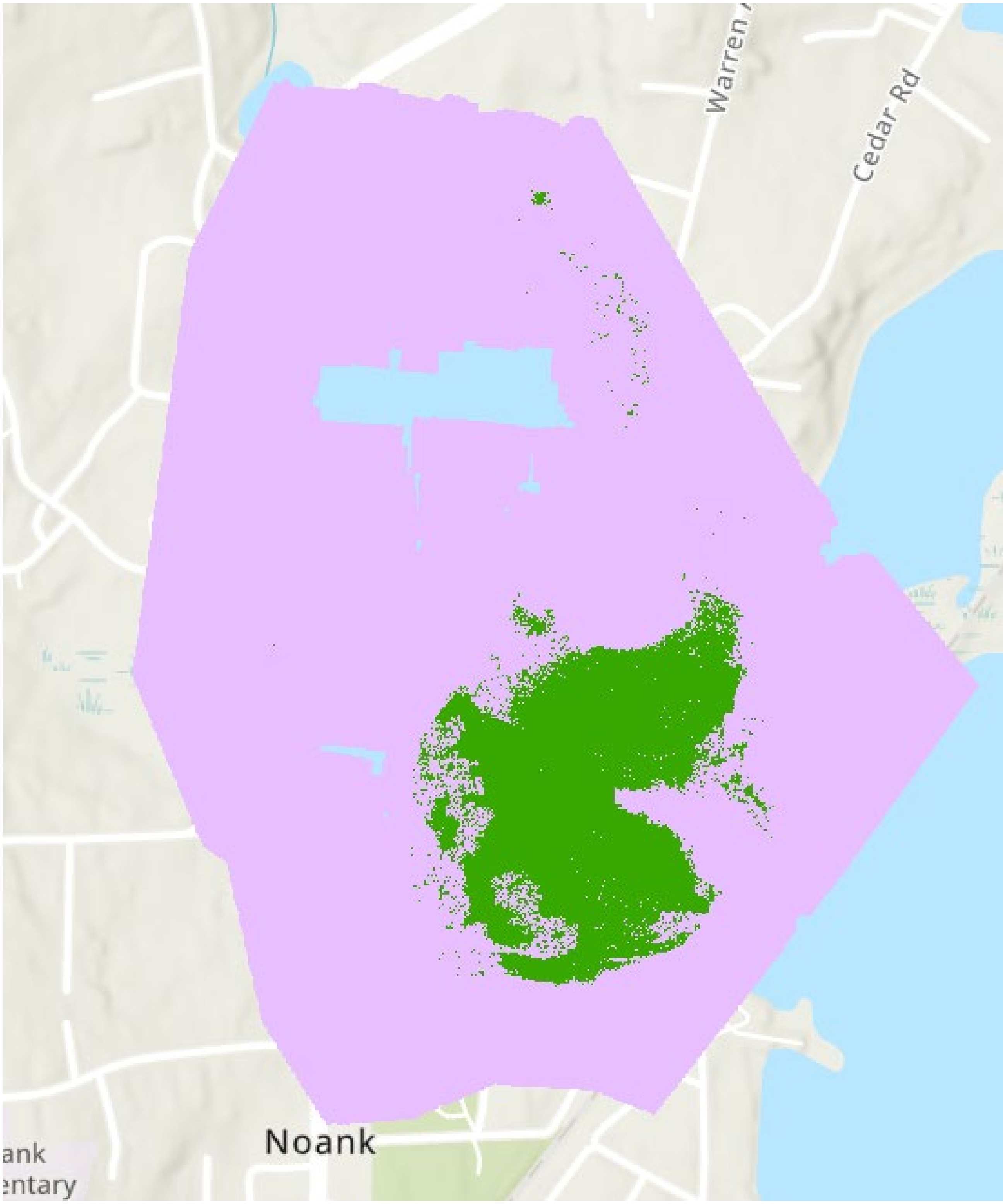


Once the raster is clipped, the image (Classified_Image_Max_Like_Reclass_Clip) will auto loaded into the ArcGIS map. The example screenshot below shows the result of the Clip Raster function and can be used for comparison to your output image. If the symbology is not as shown, adjust it as follows:

1. In the Contents pane, **Right-click** on the Classified_Image_Max_Like_Reclass_Clip raster and **Select** "Symbology". The Symbology pane will open.
2. In the Primary symbology drop down, **Select** "Unique Values".
3. In the "Field 1" drop down, **Select** Value.
4. In the Values tab, **Select** the color block next to "Eelgrass". **Click on** "Color Properties..."

and in the HEX # box **Enter** 38A800. **Click** “OK”.

5. In the Values tab, **Select** the color block next to “Other”. **Click on** “Color Properties...” and in the HEX # box **Enter** E8BEFF. **Click** “OK”.


Warren
Cedar Rd
Noank
ank
entary

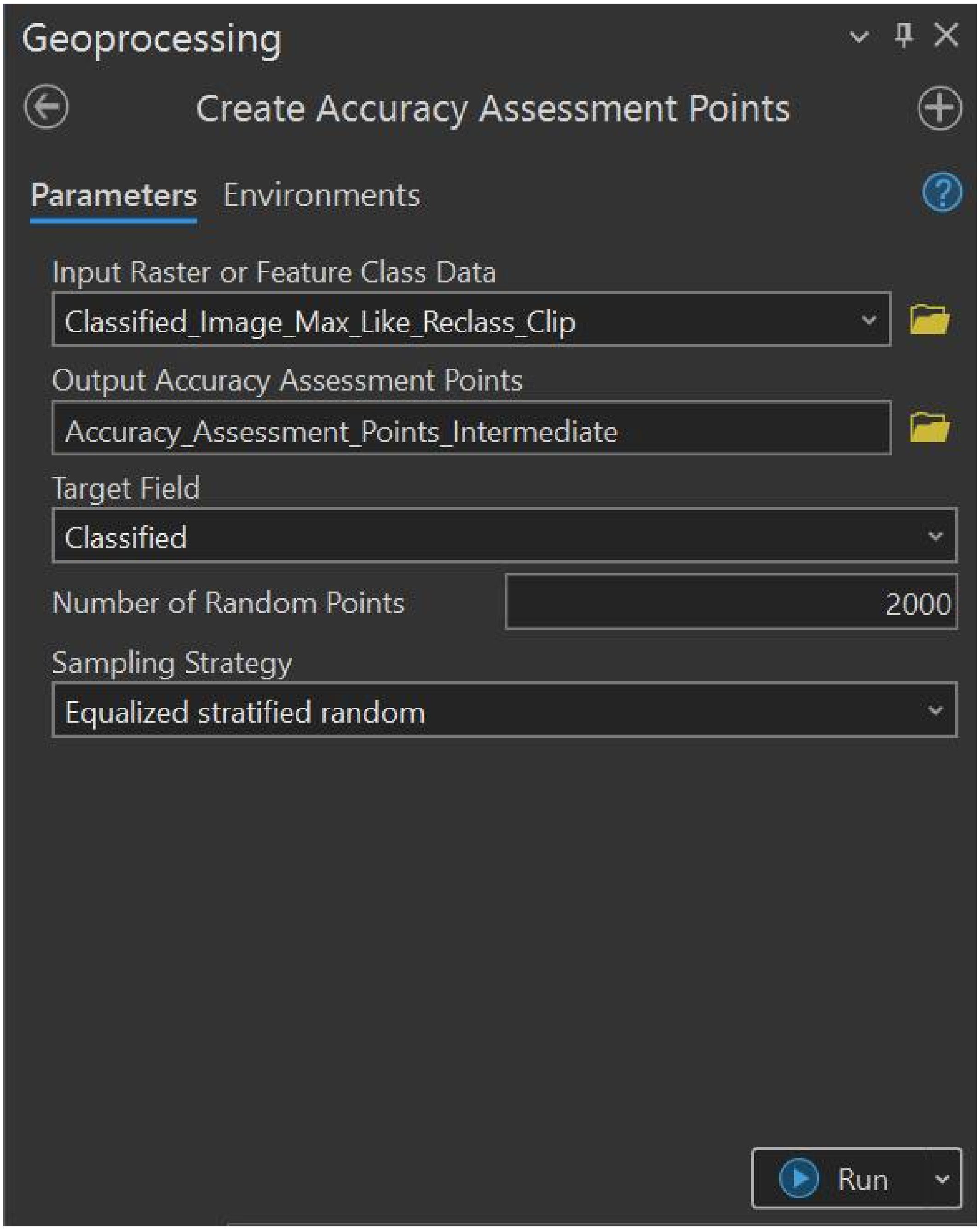
5.5.3 Complete Accuracy Assessment

Now that the reference data and classified data are in the same format and you have clipped the reclassified raster to the project boundary, we can proceed with the accuracy assessment. There will be two files created here, one being an intermediate file, and the other being the final.

1. **Click-on** the Geoprocessing Tools box in the Analysis Ribbon. The Geoprocessing pane will open. **Search** for the "Create Accuracy Assessment Points" tool. **Select** the Create Accuracy Assessment Points (Spatial Analyst Tools) geoprocessing tool.

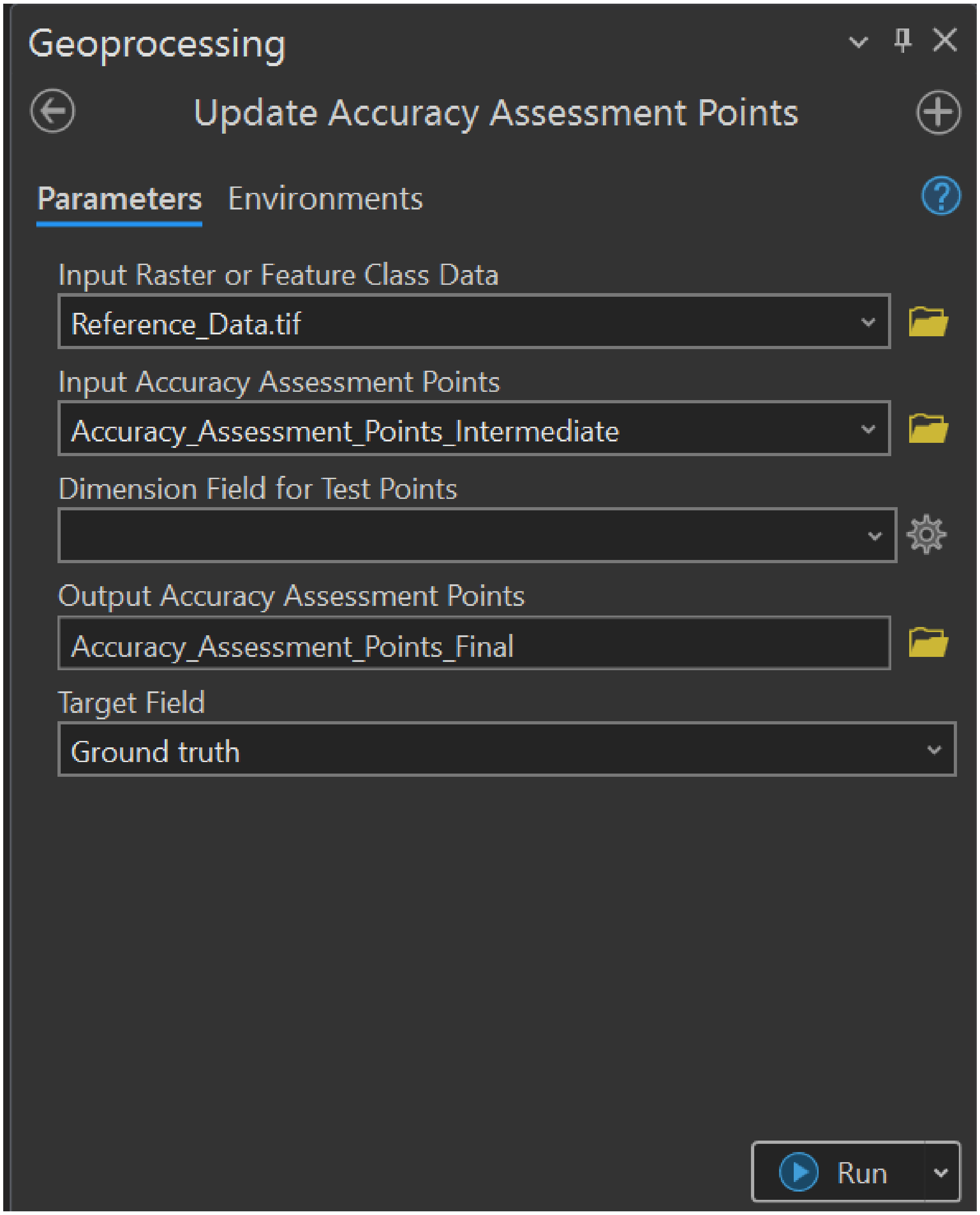
    a. For "Input Raster", **Select** the Classified_Image_Max_Like_Reclass_Clip raster from the drop-down menu.
    b. **Name** the "Output Accuracy Assessment Points" Accuracy_Assessment_Points_Intermediate and **Save** to the default project folder.
    c. For the "Target Field", **Select** "Classified".
    d. In the "Sample Strategy", **Select** "Equalized stratified random".
    e. For the "Number of Random Points", **Enter** 2,000.
    f. **Confirm** the settings with the screenshot below and **Click** the "Run" button.

Note: These points are being created to match the extent of the classified image, and then will be compared with the reference image. This action creates the intermediate points shapefile and populates the "Classified" field in the attribute table.


Geoprocessing
Create Accuracy Assessment Points
Parameters
Environments
Input Raster or Feature Class Data
Classified_Image_Max_Like_Reclass_Clip
Output Accuracy Assessment Points
Accuracy_Assessment_Points_Intermediate
Target Field
Classified
Number of Random Points
2000
Sampling Strategy
Equalized stratified random
Run

2. Also, in the Geoprocessing pane **Search** for the "Update Accuracy Assessment Points" tool. **Select** the Update Accuracy Assessment Points (Spatial Analyst Tools) geoprocessing tool.

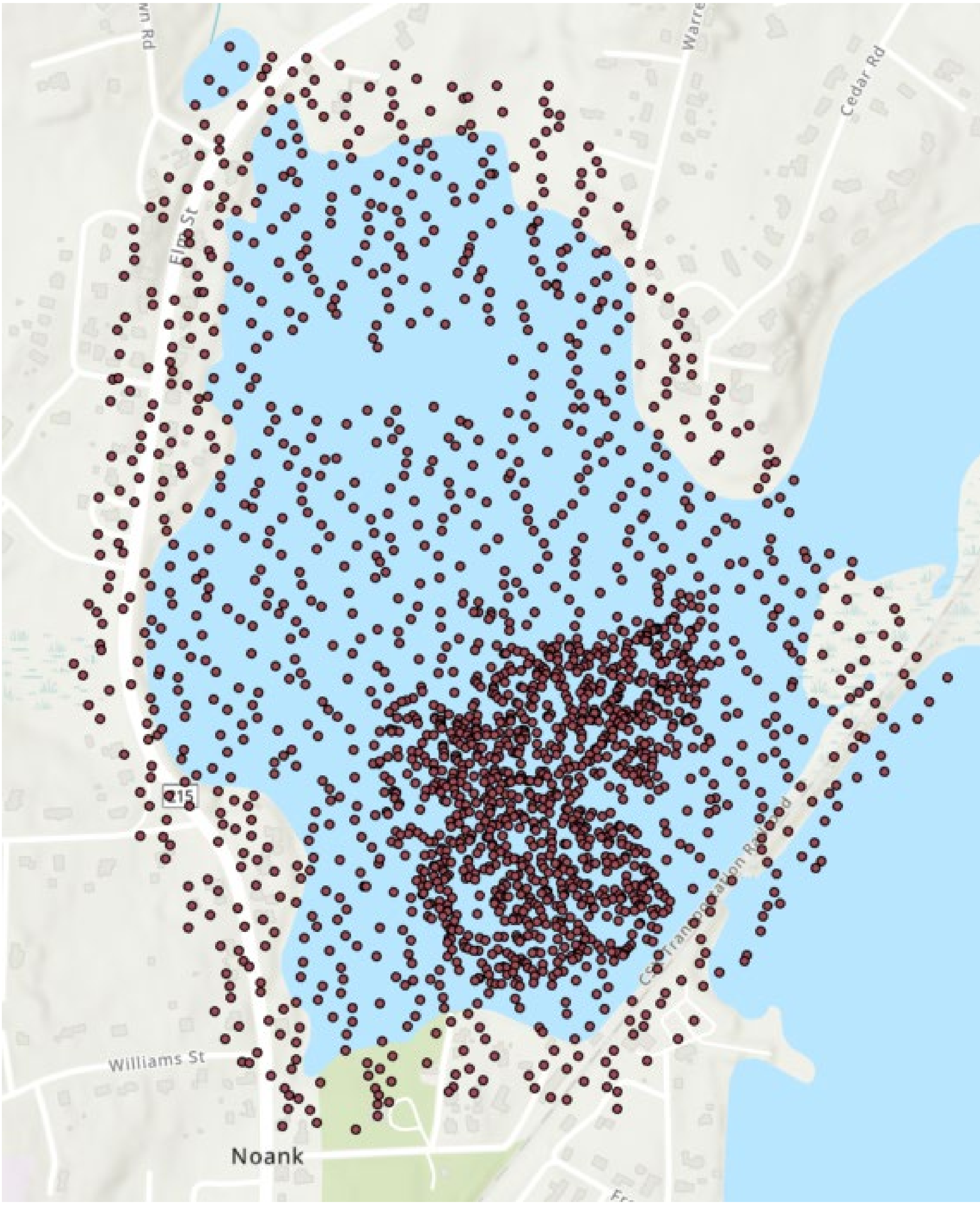
    a. For "Input Raster", **Navigate** to the Reference_Data.tif raster in the Step_0_Input_Data folder and add it to the geoprocessing tool.
    b. **Select** Accuracy_Assessment_Points_intermediate from the dropdown menu under "Input Accuracy Assessment Points".
    c. **Name** the "Output Accuracy Assessment Points" Accuracy_Assessment_Points_Final and **Save** to the default project geodatabase.
    d. For "Target Field", **Select** "Ground truth".
    e. **Confirm** the settings with the screenshot below and **Click** the "Run" button.



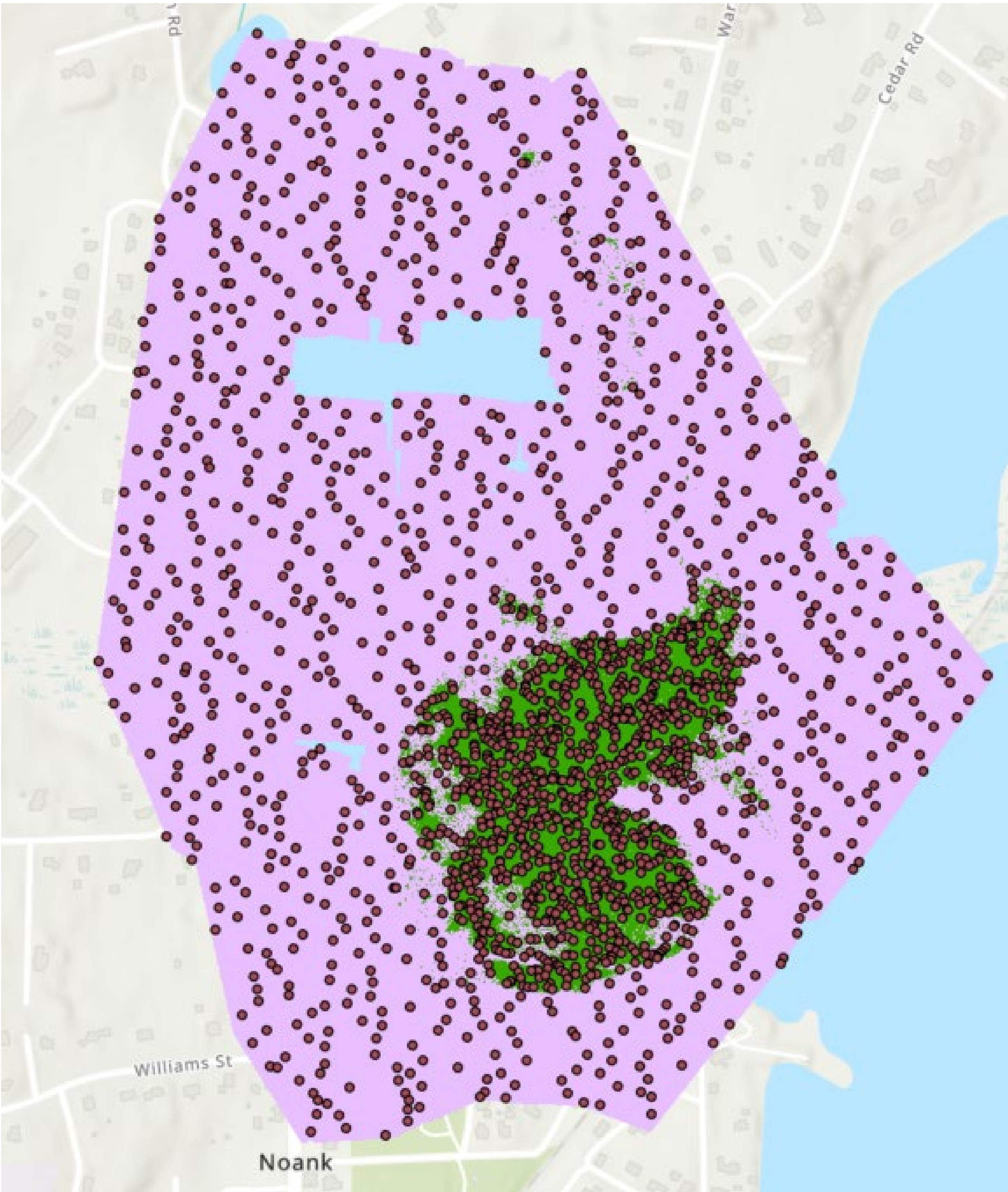
This action creates the finalized points shapefile and populates the "GrndTruth" field in the attribute table.

See the following screenshot for the points overlaid on the basemap (first image) and then overlaid on the clipped Maximum Likelihood classification (second image).

Note: The points you create may differ from what is shown in the screenshot, as there is an element of randomness inherent to the method in which the points are created in this exercise.

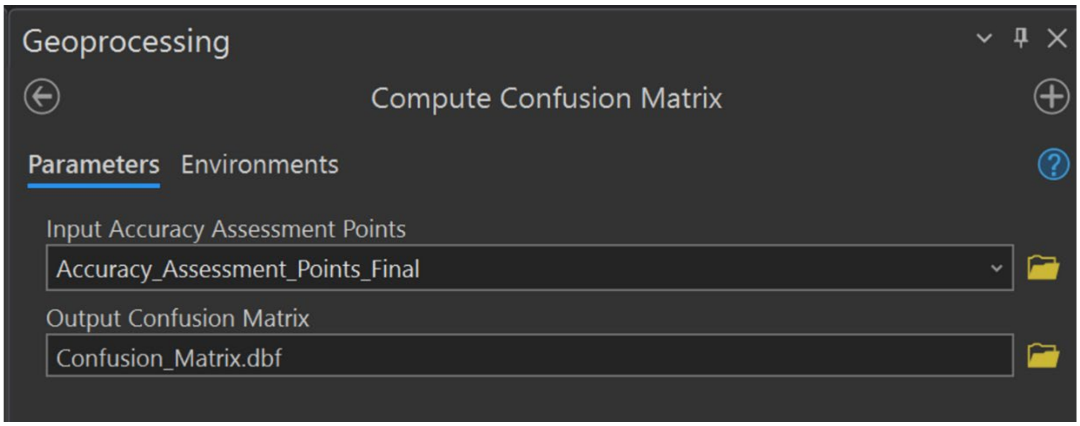


Elm St
Cedar Rd
215
Williams St
Noank
CSX Transportation Railroad

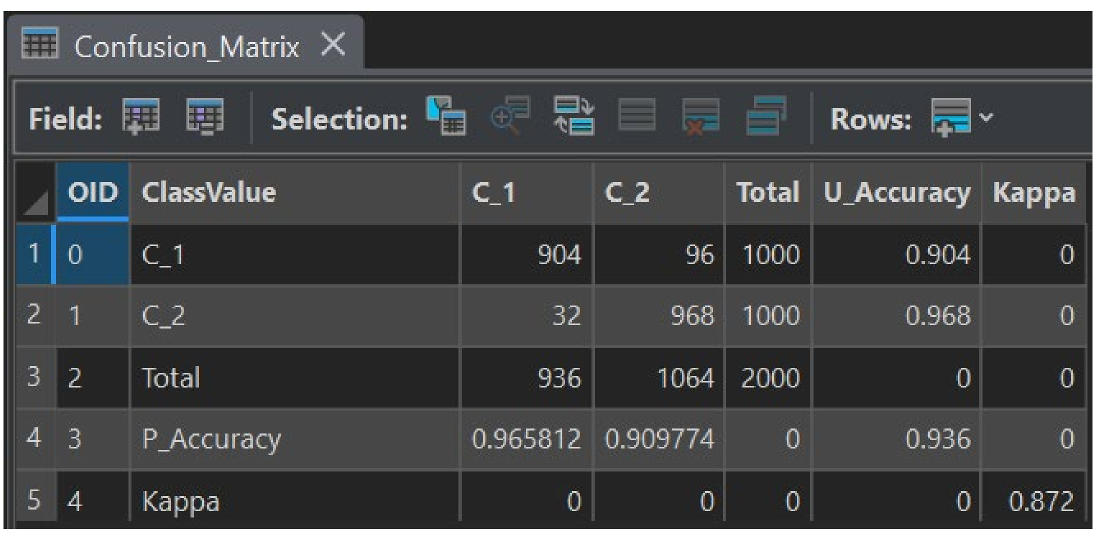

Rd
Cedar Rd
Williams St
Noank

3. Also, in the Geoprocessing pane **Search** for the "Compute Confusion Matrix" tool. **Select** the Compute Confusion Matrix (Spatial Analyst Tools) geoprocessing tool.

    a. For "Input Accuracy Assessment Points", **Select** Accuracy_Assessment_Points_Final from the drop-down menu.
    b. **Name** the "Output Confusion Matrix" Confusion_Matrix.dbf and Save to the default project folder. (Make sure the output file extension is **.dbf**, indicating a database file.)
    c. **Click** the "Run" button once all parameters are configured.



4. After generating the Confusion Matrix, **Locate** it in the "Standalone Tables" section of the Main Contents Pane, **Right-click** on it, and **Select** "Open" to see the accuracy numbers.



	OID	ClassValue	C_1	C_2	Total	U_Accuracy	Kappa
1	0	C_1	904	96	1000	0.904	0
2	1	C_2	32	968	1000	0.968	0
3	2	Total	936	1064	2000	0	0
4	3	P_Accuracy	0.965812	0.909774	0	0.936	0
5	4	Kappa	0	0	0	0	0.872

The confusion matrix (Figure 5.5a) lists the user's accuracy (U_Accuracy column) and producer's accuracy (P_Accuracy column) for Eelgrass (C-1) and Other (C_2), as well as an overall kappa statistic index of agreement. These accuracy rates range from 0 to 1 in which 1 represents 100 percent accuracy. The U_Accuracy indicates that for the total 1,000 points that the classifier used to determine eelgrass, both the classifier and reference data agreed on 904 of those points being Eelgrass. For comparison, the reference data categorized 96 points as Other, when the classifier incorrectly identified 32 Other points as Eelgrass. The Kappa statistic of agreement provides an overall assessment of the accuracy of the classification which is this case is 0.872.

The confusion matrix by itself does not provide the full context of the accuracy for the classification method. The accuracy assessment should be looked at in conjunction with a qualitative review of the output to determine the full picture. For example, the accuracy numbers will not provide any information about the quality of the classification at the edges of the eelgrass features, where the water features begin (a complex area to map manually). In addition, the accuracy numbers will change if the area is clipped to different boundaries (could be more or less accurate). The accuracy numbers could also change if a different sampling approach is used (stratified random vs. equalized stratified random) when creating the accuracy assessment points. This is to say quantitative accuracy should not be taken as an absolute measure of correctness and should be interpreted within the context of the project.

Note: The accuracy metrics may differ slightly from what is displayed in the above image, since the generated accuracy assessment points may be different than what is shown in this exercise. See this reference for an explanation of the confusion matrix.

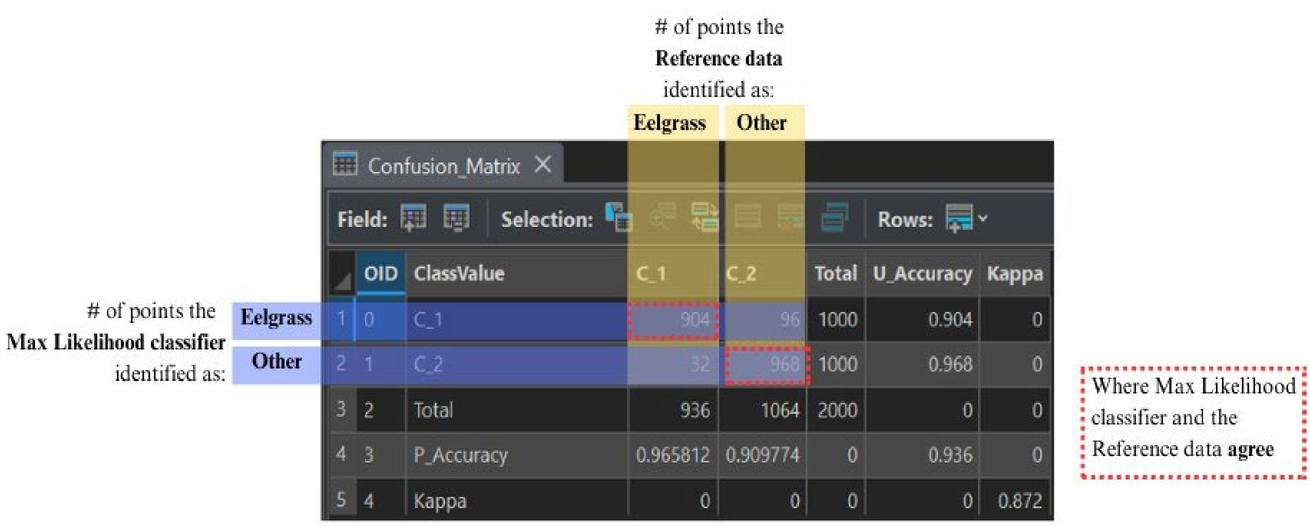
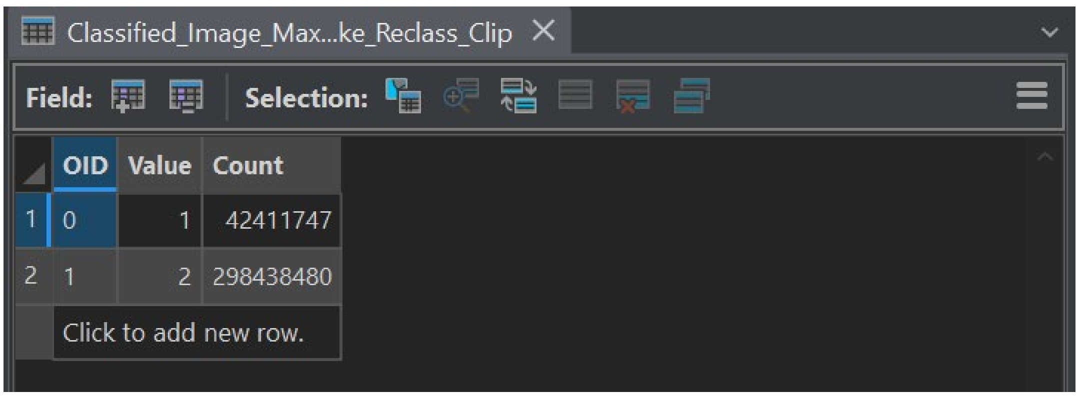


Figure 5.5a – Confusion matrix with the Reference data and Classifier results for Eelgrass (C_1) and Other (C_2).

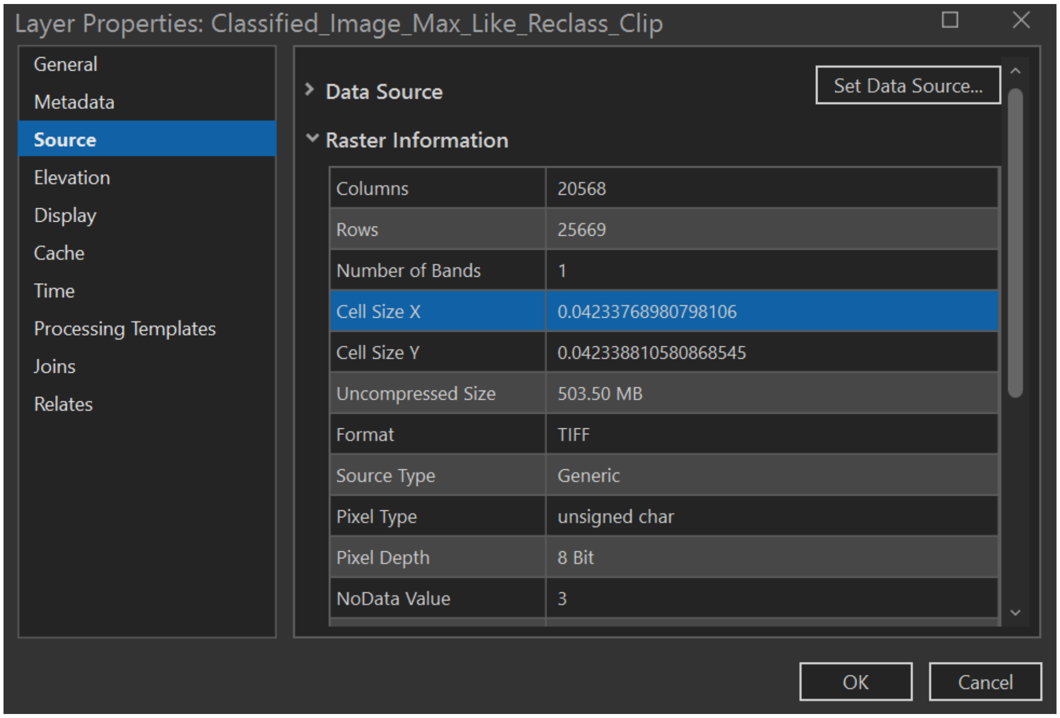
5.6 Area Calculation (5-10 minutes)

To calculate the area of features in a raster, in this case the eelgrass area, the process involves the multiplication of the number of pixels in a specific class by the area of each pixel. The number of pixels classified as Eelgrass (numeric value of 1) can be found in the attribute table of the classified image.

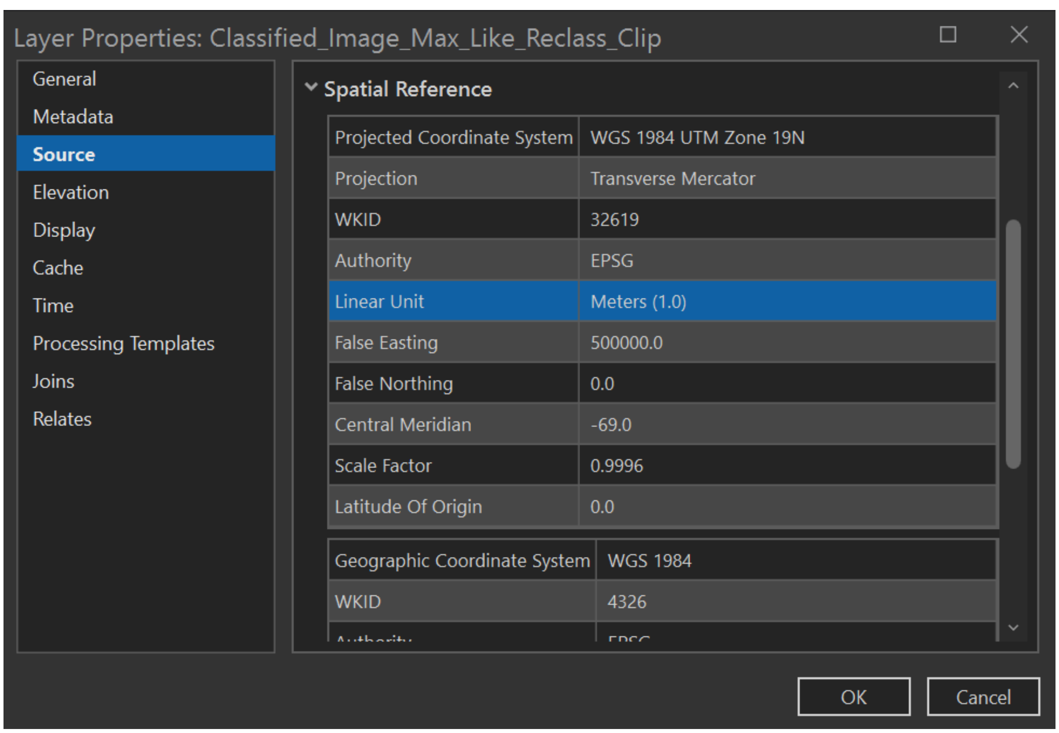
1. Open the attribute table:
    a. **Right-click** on Classified_Image_Max_Like_Reclass_Clip and **Select** "Attribute Table".
    b. **Record** the number of pixels classified with a value of 1 (see screenshot). Note that in the screenshot the "Value" column labeled as "1" represents Eelgrass.



2. Determine the area for each pixel:
    a. Open the "Raster Information" tab: **Right-click** on Classified_Image_Max_Like_Reclass_Clip and **Select** "Properties", **Click-on** "Source", then **Open** the "Raster Information" tab.
    b. **Record** the values for "Cell Size X" and "Cell Size Y".



c. **Open** the "Spatial Reference" tab (also located in the "Source" view of the layer properties).
d. **Record** the value for "Linear Unit".



3. Calculate the area of eelgrass using the following formula:

   Number of pixels * (Cell Size X * Cell Size Y).

   42411747 * (0.0423376890798106 * 0.04233881058086845)

   = 76,024.22 square meters

   = 18.79 acres

4. Compare classified eelgrass area with reference data:

   a. **Repeat** steps 1-2 above using the Reference_data raster stored in the Step_0 Input Data folder to obtain the Number of pixels, Cell Size X, Cell Size Y, and Linear Unit.

   b. **Calculate** the area of eelgrass using the formula from Step 3.

      47166633 * (0.138907694995986 * 0.1389076949959906)

      = 910,096.58 square feet

      = 20.89 acres

The area of eelgrass machine classified (18.79 acres) compares to the reference data which is based on hand drawn observation of eelgrass from imagery (20.89 acres). The maximum likelihood classification method underestimated the area of eelgrass when compared to the hand delineation outcome in the reference data. See Figure 5.6a and b for comparison.

It should be noted that direct comparison of area does not provide the full context for which measurement is more accurate. The area comparison, in conjunction with the confusion matrix and a qualitative review of the classification, should be used to determine method accuracy.

Figure 5.6 - The area machine classified as eelgrass (a) and the hand delineated reference data (b).

5.7 Discussion and Next Steps

If proceeding with machine learning classifications, there are a few adjustments that can be made to try to improve the method results. The first step is to assess the accuracy, both qualitatively (visual inspection) and quantitatively (confusion matrix, as shown) for select classification methods, and then refine the input segmentation and/or classification parameters. Default settings were used in this case, but adjustments can be made with visual comparisons to observe for improvements in the outcomes. The segmentation or classification documentation can provide additional information on this. It may also be helpful to create additional training samples, which will provide more information to the classification models and likely improve the accuracy.

It's important to note that e-Cognition software was initially evaluated for this project. However, it was found that the software rounded the spectral data when loading raster bands. This resulted in a loss of the details needed to differentiate features making the data useless in computing spectral indices. Considering this and the fact that CT DEEP has user licenses for ESRI's product, the decision was made to proceed with the classification analyses in ArcGIS Pro.

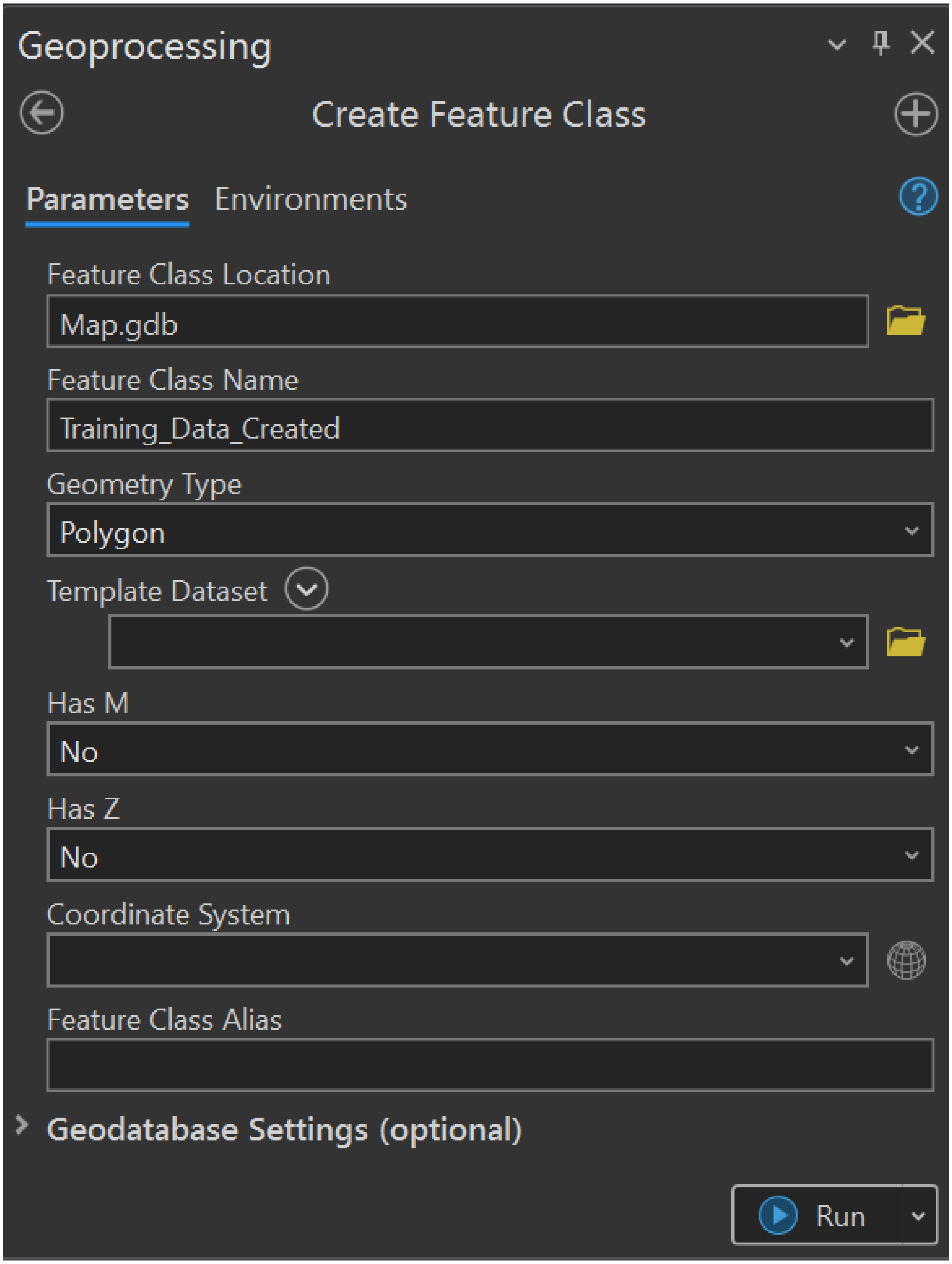
While this method is useful and can be applied to other dataset, it's important to note that advancements in technology have occurred since this project was initiated in 2021. For example, deep machine learning with AI in a neural network has been further developed with expanded use. It is recommended that those interested in pursuing an automated approach to eelgrass delineation, evaluate the various available technologies specific to their project. Well defined objectives and long-term needs should guide the review and final technology selection.

Appendix A: Create Training Samples without Field Data

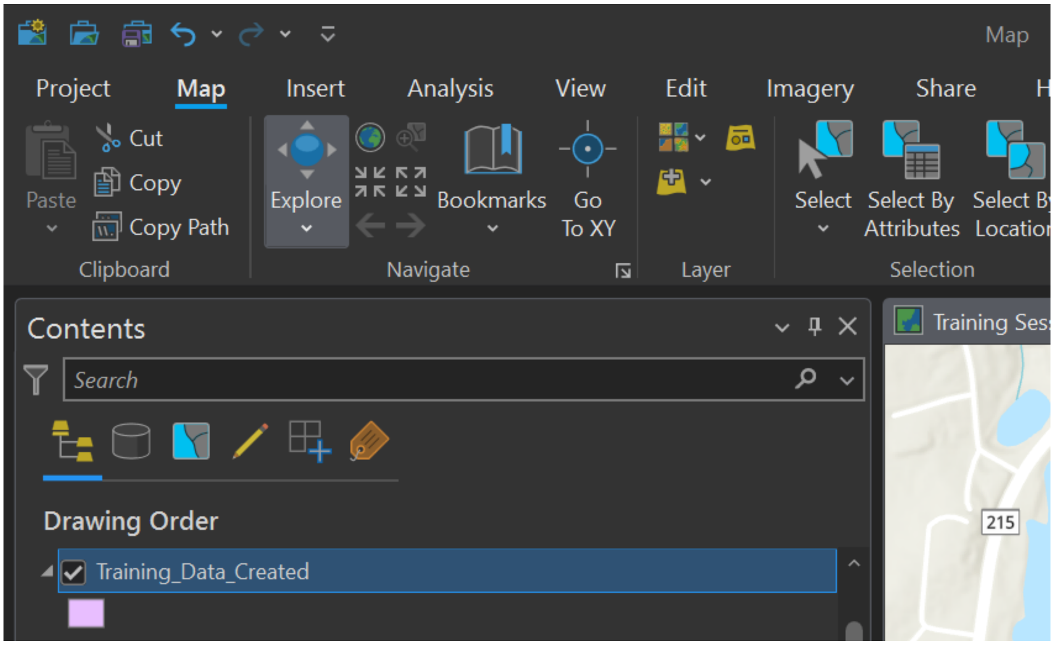
This section describes how to create training samples if field data is not already available.

1. In the Contents pane, **Right-click** on Multispectral_Imagery.tif and **Select** “Zoom to Layer”.
2. Adjust the image symbology so the features for classification (i.e. Eelgrass and Other) are visually clear:
    a. In the Contents pane, **Right-click** on the layer and Select “Symbology”. The Symbology pane will open.
    b. **Set** the “Primary symbology” to RGB, “Red” to NIR, “Green” to Red, and “Blue” to Green.
    c. In the “Stretch type”, **Select** Histogram Equalize.
    d. In the “Statistics” tab**, Select** Dataset.
3. **Open** the “Create Feature Class” geoprocessing tool to create a shapefile. **Save** the Feature class to the default geodatabase and **Name it** “Training_Data_Created. **Select** the Polygon Geometry Type. **Confirm** settings with the screen shot below and **Select** Run.

   Note: The “Point” can be selected as the Geometry type. If this option is used the points may need to be buffered by some distance to include enough pixels, though the same effect can be achieved with the method demonstrated here. Both polygons and buffered points must contain enough pixels to inform the classification process.

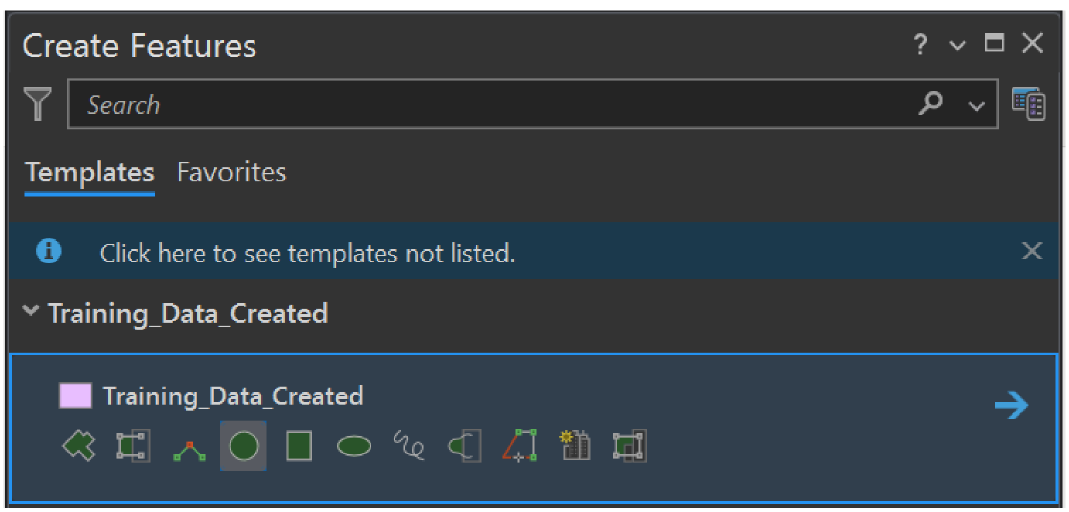


4. Create polygons (circles or rectangles) within the shapefile representing the features: In the Contents pane, **Right-click** on "Training_Data_Created". **Select** the "Edit" tab/ribbon. **Select** "Create" (within the **Features** window). This opens the **Create Features** window on the right side of the program.

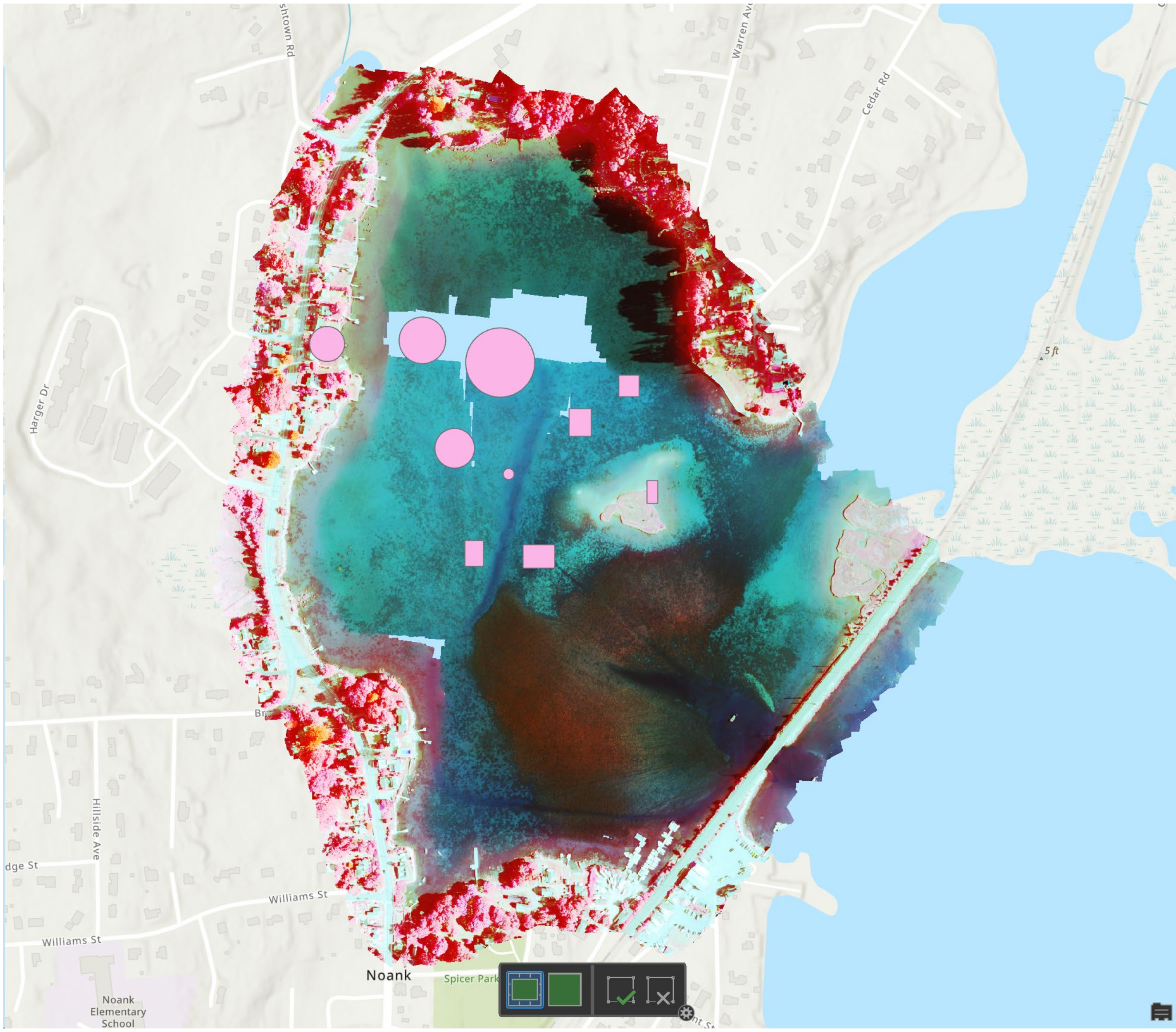


e. Within the “Create Features” window, **Select** the Training_Data_Created file. **Choose** from the circle or rectangle options to delineate your samples.

Note: The default tool (Create a polygon feature) option is not best option for this exercise. Select the circle of rectangle tool. You can also use a combination of both (see later screenshot of example shapes).



f. Draw features (circles or rectangles) representing the Eelgrass and Other (see screenshot). The screen shot shows polygons for example purposes only. Do not use samples as large for this project. This is just showing the difference types of polygons you can create during the process. See the following text and screenshots for additional context about the samples.



Note: Training samples should be representative of the features within the study area. For smaller projects like Beebe Cove (~130 acres), at least 50 sample points or polygons for each class should be created. For the Beebe Cove example, that would be Eelgrass (50 samples) and Other (50 samples) with a final count of 100 training samples. These samples should span the spatial extent *of the imagery, as to provide a good representation of features within the full area. Open water with no apparent eelgrass should be labeled "Other". Additionally, if you have features like docks, boats, and land features such as impervious surfaces and trees, include these as samples of "Other". Samples should be large enough to cover a useful number of pixels so that the feature it represents is obvious visually and contain enough pixels to provide enough spectral information, but at the same time, be small enough so that each sample only contains a single feature (Eelgrass or Other, not partial of each). Samples that contain more than one feature will propagate errors through the classification and degrade the output results. The following figures demonstrate optimal and suboptimal training sample creation.*

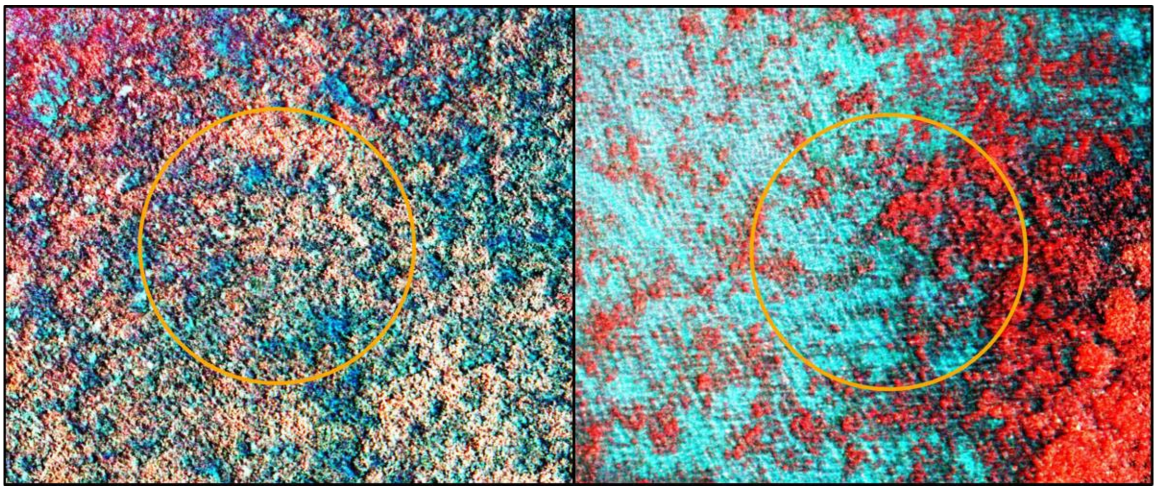


Figure a. Screenshots of an optimal and suboptimal sample, both at 1:125 scale. The left sample is optimal because it contains 90% or more eelgrass, while the right sample is suboptimal because it contains 30-40% eelgrass and 60-70% water, which would likely confuse the classification methods and result in erroneous output.

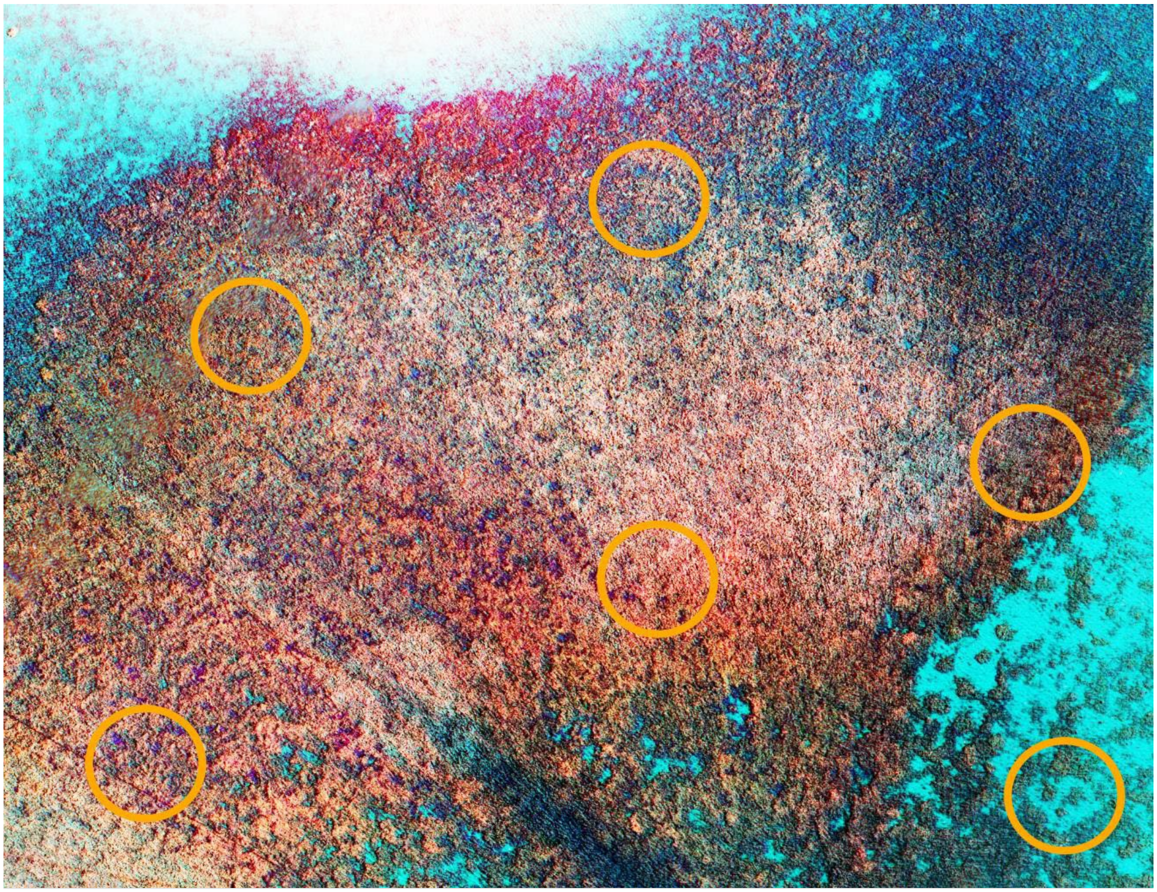


Figure b. Screenshot of five optimal samples and one suboptimal sample (southeast corner of the image), at a 1:600 scale. Samples near the edge where eelgrass ends and water begins can be the most difficult, and those areas may need to be avoided when creating training samples.

Once satisfied with the polygons. Close the “Create Features” tool and continue to add the necessary fields required for classification.

g. **Right-Click** on Training_Data_Categorized.shp in the Contents pane and **Open** the Attribute Table.

   Insert a column for classname: In the Attribute Table, **Click-on** “Add” next to “Field:” The “Fields Tab” will open. **Type** in “classname” in both the “Field Name” and “Alias” columns and **Select** “Text” from the “Data Type” drop down menu. **Close** the “*Fields: Training_Data_Categorized” tab. **Save** the changes and **Confirm** the new “classname” column in the Attribute Table.

   Insert a column for classvalue: In the Attribute Table, **Click-on** “Add” next to “Field:” The “Fields Tab” will open. **Type** in “classvalue” in both the “Field Name” and “Alias” columns and **Select** “Long” from the “Data Type” drop down menu. **Close** the “*Fields: Training_Data_Categorized” tab. **Save** the changes and **Confirm** the new “classvalue” column in the Attribute Table.

h. Enter the information related to the class name and class value for all the training samples. This data will be entered manually.

i. In the “Training_Data_Categorized Attribute Table:

   Right-click on the “classname” field and **Select** “Calculate Field”. This opens the “Calculate Field” window. See the below screenshot for input details. **Click** “Apply” to populate the field, then “OK” to exit the window.

   Next, **Right-click** on the “classvalue” field and **Select** “Calculate Field”. This opens the “Calculate Field” window. See the below screenshot for input details. **Click** “Apply” to populate the field, then “OK” to exit the window.

Once the classname and classvalue fields are populated with the correct information, the shapefile can be used in the classification. Go to Section 5.3 to continue Image Segmentation.