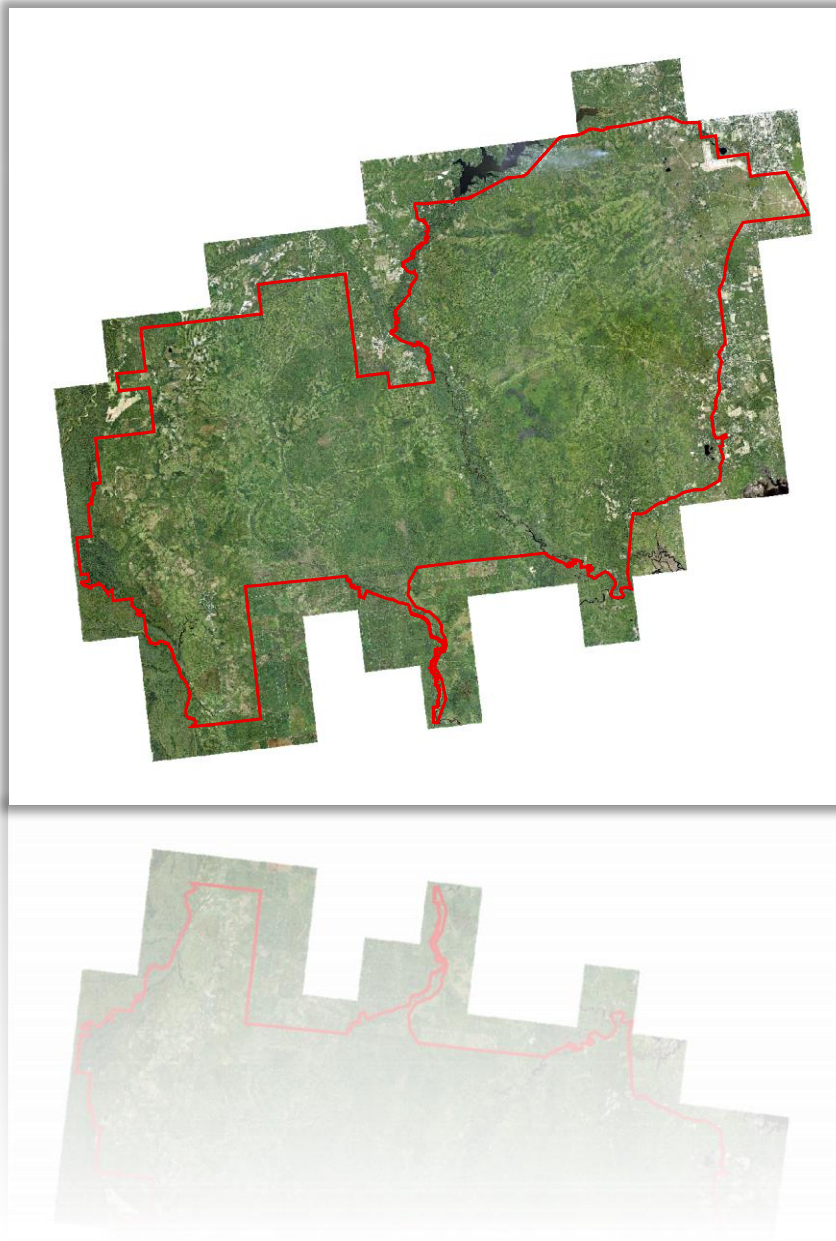


IMAGE BASED CLASSIFICATION USING THE RMRS RASTER UTILITY TOOLBAR: FOCUS ON WORKFLOW

JOHN HOGLAND, KORYN HAIGHT & JOE ST. PETER



CREATING A CLASSIFICATION

OVERVIEW

In this workshop, we are going to create a polygon based classification from readily available vector and raster datasets. Topics covered will include raster processing, sampling, classification, and model development and interpretation. Each section in this tutorial walks the reader through one aspect of making a classification and is designed to be worked through in order.

Section 1: [Setup and Background](#). This section describes which tools to load into ArcMap and the datasets that will be used throughout the tutorial.

Section 2: [Raster Processing](#). This section walks users through how to create predictive variables from readily available raster datasets.

Section 3: [Sampling](#). This section describes how to determine an appropriate sample size and how to select samples from the population.

Section 4: [Classification](#). This section walks users through how to build and apply a classification model using random forest classifier.

Section 5: [Interpretation](#). This section walks users through interpreting the classification model and the associated outputs.

All users should read the “[Setup and Background](#)” section to better understand the datasets being used within these exercises. From start to finish this tutorial will take approximately 3 hours to complete.

REQUIREMENTS AND DATA LOCATION

The two requirements to complete this tutorial are ArcGIS 10.0 (or higher) and the zip file titled ‘ClassificationModeling.zip’. The zip file contains this document, a data folder, and a model folder. The “data” folder contains all datasets needed to work through the tutorial. The “model” folder is empty and will be used to store models created while working through the tutorial. It should also be noted that, if you do not already have it, you must download and install Microsoft .Net Chart tools 3.5 sp1 in order to build the graphics in this tutorial.

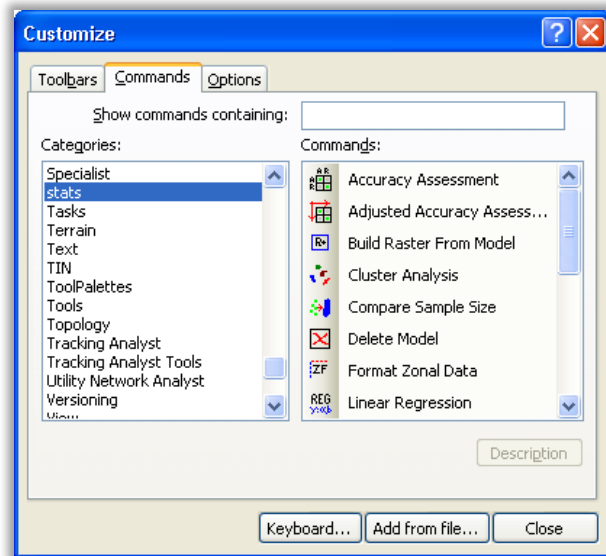
SECTION 1: SETUP AND BACKGROUND

If you have not already done so, please download and install the RMRS raster utility toolbar from the RMRS raster utility website ([link](#)). To access many of the tools for these exercises you will need to customize ArcMap and add 10 commands to the RMRS Raster Utility Toolbar. To customize the RMRS Raster Utility Toolbar:

Right click in the toolbar area.

Left click “Customize” (at the bottom of the dropdown).

This will bring up the customization window.



From the customization window:

Select the “Commands” tab.

Scroll down to the ‘stats’ commands category.

You can add commands to a menu by left-click holding on each command, dragging that command to the menu (the menu will open when you hover over it) and releasing the left click button.

Add the following buttons from the ‘stats’ commands category to the “Statistics Modeling” menu:



“Random Forest”



“Predict New Data”



“PCA/SVD”



“Show Model Report”



“Format Zonal Data”



“Build Raster From Model”



“Accuracy Assessment”

Add the following button from the ‘sample’ commands category to the “Sample” menu:



“Select Samples from Model”

Add the following buttons from the ‘rasterUtilities’ commands category to the “Raster Analysis” menu:



“NDVI”



“Zonal Statistics”

To demonstrate how to use these tools, we will be using the data within the data folder. Within the data folder there is a raster and vector dataset. The raster dataset, NAIP_30.img, was resampled from NAIP imagery and represents the base surface that will be transformed into predictive variables and summarized for the

classification. The vector datasets are segmented polygons ("Segments.shp") that represents the population of polygons we will classify and use to evaluate our classification.

SECTION 2: BUILDING PREDICTIVE VARIABLES (RASTER PROCESSING)

The RMRS Raster Utility toolbar has many tools that can be used to manipulate, summarize, and sample raster values¹. In this exercise, we will create a series of summarized predictive surfaces that will be used to classify land cover categories for segmented polygons. The predictive variables we will create, summarize, and use include: outputs of a principal component analysis (PCA), and a standard deviation for a 3 by 3 moving window of the PCA outputs. The NAIP_30 dataset is a 30 meter resampled raster dataset derived from 1 meter false and color infrared NAIP imagery².

To begin with, let's

Open a blank document in ArcMap and load Segments and NAIP_30 into the map. Segments and NAIP_30 can be found within the data folder.

Often, imagery such as the NAIP has highly correlated data across spectral bands. To remove this correlation and potentially reduce the number of predictive classification variables we will perform a PCA:

Open the "Statistical Modeling" dropdown menu.

Left click the "PCA/SVD" button (*).

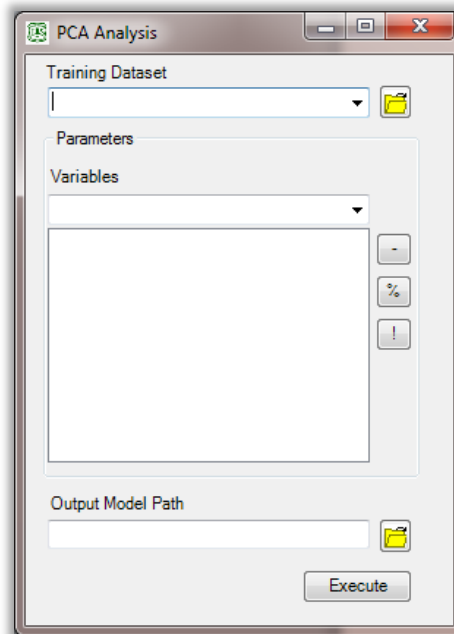
The "PCA Analysis" form will open.

Open the "Training Dataset" dropdown.

¹ For more information on RMRS Raster Utility tools that can be used to manipulate, summarize, and sample raster datasets please download the function modeling tutorial from the RMRS Raster Utility website ([link](#)).

² Link to NAIP program <http://www.fsa.usda.gov/FSA/apfoapp?area=home&subject=prog&topic=nai> and downloads <http://datagateway.nrcs.usda.gov/GDGOrder.aspx>.

Select the “NAIP_30” dataset.



Notice that the Variables selection box is now grayed out. This means that all bands within the NAIP_30 raster dataset will be used in the PCA.

Click on the folder icon next to the “Output Model Path” text box.

Navigate to the ~\Model folder.

Name the PCA model file “NaipPcaModel”.

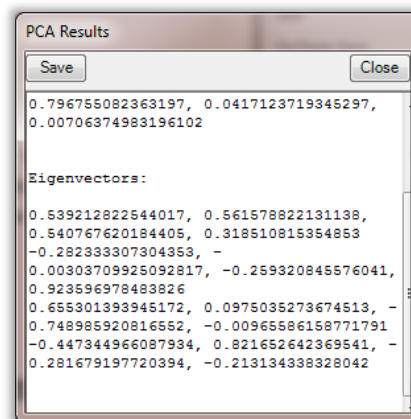
Click “Save”.

Left click “the Execute” button.

Please be patient while the analysis is running.

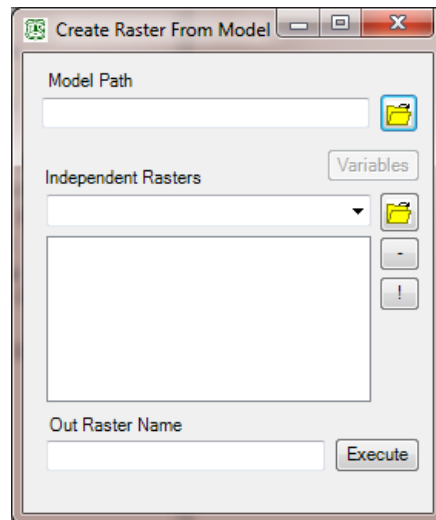
When completed, the PCA\SVD tool will produce a PCA model and report that can be used to generate a new PCA multiband raster surface.

Next, let’s create PCA raster surface using the “NAIP_30” raster dataset and our newly created “NaipPcaModel.mdl” file.



Open the “Statistical Modeling” dropdown.

Left click the “Build Raster From Model” button () located within menu.



Click on the folder next to the “Model Path” text box.

Navigate to and select the “NaipPcaModel.mdl” file.

Open the “Independent Rasters” drop down.

Select each of the bands of the NAIP_30 raster dataset in ascending order (the default order).

Type in PCA_30 in the “Out Raster Name”.

Click “Execute”.

This action will create a new temporary function raster dataset called PCA_30 and add it to ArcMap’s table of contents³. Next let’s convert the temporary dataset to a permanent dataset.

Click save button () on the RMRS utility toolbar.

Under “Raster to Save”, select “PCA_30”.

Click on the folder icon next to the “Out Workspace”.

Navigate to and highlight “data” workspace.

Click “Add”.

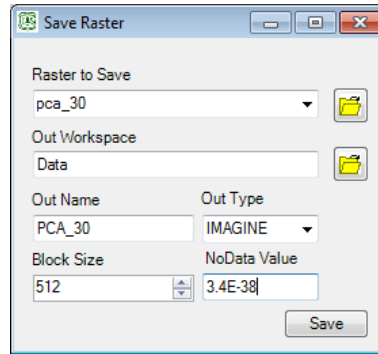
Name the dataset “PCA_30”.

Set Out Type to “IMAGINE”

Leave the “Block Size” at default.

Leave the NoData Value at the default

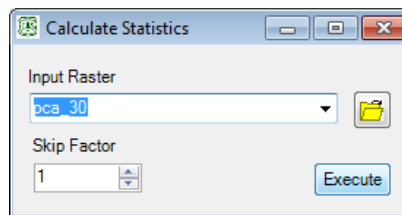
³ Depending on the version of ArcMap the display can be completely black. This is due to the statistics for the dataset. Newer versions of ArcMap auto calculated statistics when datasets are added to the table of contents while older versions do not.



Click “Save”.

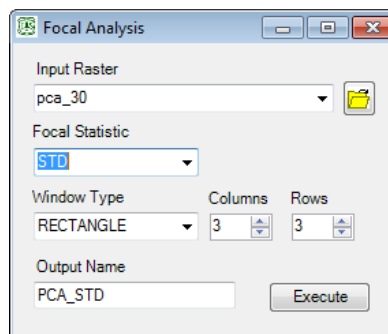
Note, upon displaying the newly created raster dataset the display appears gray. The reason for this is that raster band statistics have not been calculated. To fix the display:

Left click on the “Calculate Statistics” button () in the RMRS utility toolbar.
 Select the “PCA_30” raster.
 Choose a “Skip Factor” of 1 cell.
 Click “Execute”.



Next let’s make a first order texture surface of the “PCA_30” raster dataset.

Open the “Raster Analysis” menu.
 Click the “Focal Analysis” button ().
 For “Input Raster”, select “PCA_30”.
 For “Focal Statistics”, select “STD”.
 For “Window Type”, select “Rectangle”.
 Specify a window size of 3 columns by 3 rows.
 Name the output “PCA_STD”.



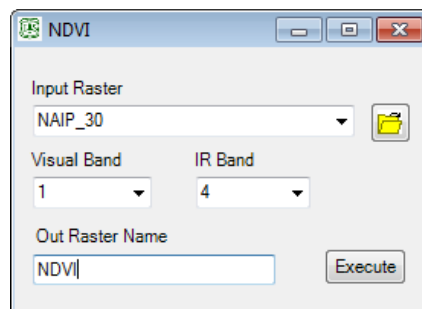
Click “Execute”.

Convert the temporary "PCA_STD" to a permanent raster dataset just as we did with PCA_30:

Click the "Save" button () on the RMRS utility toolbar.
Under "Raster to Save", select "PCA_STD".
Click on the folder icon next to the "Out Workspace".
Navigate to and highlight the 'data' workspace.
Click "Add".
Name the dataset "PCA_STD".
Click "Save".

Next let's create an NDVI raster surface using the "NDVI" tool.

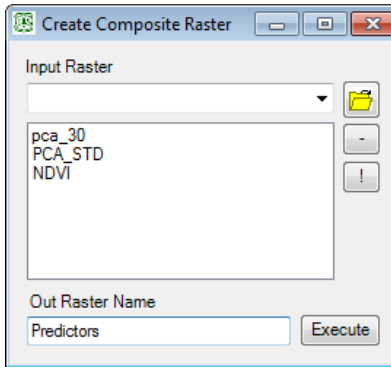
Open the "Raster Analysis" menu.
Click the "NDVI" button ().
For "Input Raster", select the "NAIP_30" dataset.
For "Visual Band", select "band 1".
For "IR Band", select "band 4".
Name the output "NDVI".



Convert the temporary "NDVI" dataset to a permanent dataset called "NDVI" within the "data" workspace (just as we did with PCA_30 and PCA_STD).

Next, we will combine our predictor surfaces raster datasets into one multiband raster dataset called "Predictors".

Open the "Raster Analysis" menu.
Click the "Create Composite Raster" tool ().
Select all of the following for "Input Raster":
 PCA_30
 PCA_STD
 NDVI
For "Out Raster Name", type "Predictors".
Click "Execute".



Notice that these raster datasets only took seconds to create. This is due to the dynamic way that RMRS Raster Utility process raster type datasets ([function modeling](#)). Finally, to summarize our predictor values for each polygon we will use the “Zonal Stats” tool. Unlike ESRI’s zonal statistics tool⁴, the RMRS Raster Utility “Zonal Stats” tool will perform zonal statistics on multiband raster datasets, has fewer limitations, accounts for overlapping polygons, and will output a table with summarized values based on both the linked field and raster band.

Open the Raster Analysis menu.

Click the Zonal Stats tool ().

For the “Input Zonal Raster or Feature Class” combo box, select the “Segments” feature class.

For the “Zone Field” combo box, select OBJECTID.

For “Zonal Statistics”, select “MEAN”.

For the “Input Value Raster”, select “Predictors”.

Click on the open button () next to “Output Table Path”.

Navigate to the ‘data’ workspace.

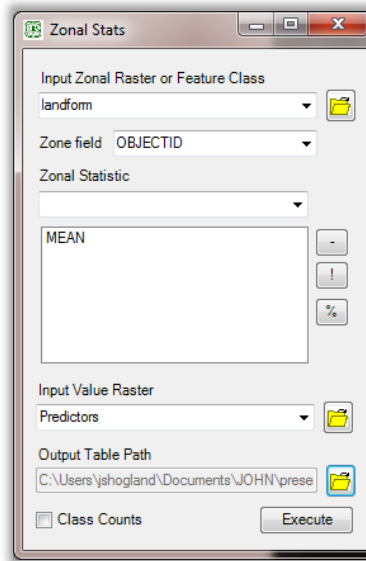
Type “zonalPred” in the text box.

Click “Save”.

Click “Execute”.

⁴ Link to description of ESRI’s zonal statistics function ([link](#))

Please be patient while the analysis is running. This one takes a little while.



This will create a new table within the “data” workspace called “zonalPred_VAT”. Now, we are going to transpose that table and join those values to the “Segment” polygon layer using the “Format Zonal Data” tool (ZF) located in the “Statistical Modeling” menu.

Open the “Statistical Modeling” menu.

Click the “Format Zonal Data” tool (ZF).

For the “Zonal Dataset Table” combo box, select the “Segments” polygon feature class.

In the “Link Field” combo box, select “OBJECTID”.

Leave the “Field Prefix (optional)” box blank.

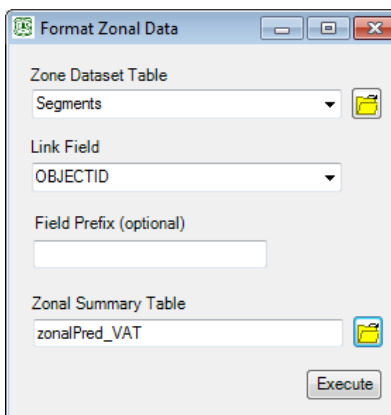
Click the open file button (F) next to “Zonal Summary Table”.

Navigate to the “zonalPred_VAT.dbf”.

Click “Add”.

Click “Execute”.

Please be patient while the analysis is running.



Upon finishing, 18 new fields will be added to the “Land cover” feature class: Count-1-9 and MEAN_1-9. The values within those new fields correspond to the summarized values in the “zonalPred_VAT” table and are now ready to be used as inputs to predict land form types.

SECTION 3: SAMPLE SIZE & SELECTION (SAMPLING)

So now that we know how to create predictive variables using function modeling and the “zonal Stats” tool, how many samples should we take across the population of polygons to build our classification model and which records should we sample? To answer these questions we must think about what characteristics we want out of our sample. The first and most important characteristic of our sample is that it is representative of the population. In addition, we would like our sample to represent our population using as few locations as possible that are relatively quick and easy to acquire. One way to do this is to use heuristic rule to estimate the total number of samples needed to train a classification model. In this exercise we will use a heuristic rule of 30 samples per land cover class.

Our land cover classes consist of easily identifiable types within the extent of the NAIP imagery (Table 1). Given that we have seven land cover types, we will need a total of 210 samples collected from our “Segments” feature class (30 coming from each land cover type).

TABLE 1. LAND COVER CLASSES AND DESCRIPTION

<u>Code</u>	<u>Label</u>	<u>Description</u>
3000	Grass	Grasslands, pastures, and other fields
4000	Deciduous	Broad leaf tree species that lose all their leaves in the winter
4500	Evergreen	Needle leaf tree species that typically do not lose their needles
5000	Water	Surface areas such as ponds, lakes, and large streams
7500	Urban	Areas composed primarily of manmade structures such as houses, roads, buildings, parking lots, and bare soil.

To select representative samples from our “Land cover” feature class, we will visually find 30 polygons that represent each of the Land cover classes and label the “LC” field based on the codes within Table 1. The assumption here is that by select 30 samples of each class we will capture the full range of that class’ variability and closely match the empirical distribution of the predictive variables. First, we’ll make the polygons in our “Segments” layer transparent so we can view the polygons in our “Segments” layer on top of our “NAIP_30 layer”.

- Right click on the “Segments” layer.
- Go to “Properties”.
- Navigate to the ‘Symbology’ tab and click ‘Features’.
- Left click on the “Symbol” box.
- Select “Hollow” (on the left).
- Change the outline color to a bright color (black would be difficult to see against the NAIP imagery.)
- Click “OK”.
- Click “OK” again.

Now, we’re going to visually find and select 30 polygons for each Land cover class to train our model.

- Turn off all layers except the “Segments” and “NAIP_30” layers.

Make sure the “Segments” layer is above the “NAIP_30” layer in the Table of Contents.
Right click on the “Segments” layer.
Select “Open Attribute Table”.

Notice that the column “LC” shows ‘0’ in each cell. We’re going to fill 30 of these cells for each class by visually finding and selecting 30 polygons for each Land cover class and assigning them the corresponding codes.

Make sure you have the “Editor” toolbar open. (If not, right click in the toolbar area and select “Editor”.)
In the “Editor” toolbar, left click the “Editor” button.
Left click “Start Editing”.
Click “OK”.
The “Create Features” box will open (if it doesn’t ‘pop up’, it may be pinned on the right).
Left click “Segments”.

We’ll start with the Land cover class “Grass”. Noted in Table 1, the description of this Land cover class is “Grasslands, Croplands, and other fields that have clear signs of irrigation”.

Left click the “Select Features” icon in the “Tool” toolbar.
Do your best to visually identify areas with signs of irrigation. Don’t fret about perfection.
Hold down the “Shift” key on the keyboard.
Select 30 polygons that visually match the description for “Grass”.
(Note: you can select multiple polygons in an area by clicking and dragging the mouse.)
Right click on the “LC” column in the “Segments” attributes table.
Select “Field Calculator”.
In the “LC =” box, type the Land cover class code (which is 3000 for “Grass”).
Click “OK”.

We just selected 30 samples representative of the Land cover class “Grass” and labelled the “LC” for these samples with the appropriate code, 3000. We’ll walk through this process again for “Water”.

Left click the “Clear Selected Features” icon in the “Tool” toolbar.
Visually identify areas that have water.
(Again, don’t fret about perfection.)
Select 30 polygons for the Land cover class “Water”.
Right click on the “LC” column in the “Segments” attributes table.
Select “Field Calculator”.
In the “LC =” box, type the Land cover class code (which is 5000 for “Water”).
Click “OK”.

Next, we’ll repeat this process for each remaining Land cover class (Deciduous, Evergreen, Urban) and label each “LC” with the appropriate code (4000, 4500, 7500 respectively). Some classes are more challenging than others. For instance, the class “Deciduous” may be more challenging to visually identify than other classes, but we can reason that deciduous trees are most likely to be the forested areas right next to streams. Don’t fret about perfection, simply do your best to visually identify polygons that fit the description for each Land cover class (refer to Table 1 for Land cover class descriptions and codes).

Repeat the above process for each remaining Land cover class, labelling the “LC” column accordingly.
(Be sure to “Clear Selected Features” between each Land cover class.)

After you have completed these steps for each Land cover class, left click on “Editor”.
Select “Stop Editing”.
It will prompt you to save.
Click “Yes”.

Upon completion, you will have built a training dataset for a classification model and can now use statistical modeling to predict “Land cover” types based on our visually selected samples. We will do this in the next section using the Random forest modeling algorithm.

SECTION 4: RANDOM FOREST LAND COVER MODEL BUILDING (CLASSIFICATION)

Random Forest is a data mining, ensemble modeling technique based on multiple classification or regression trees⁵. This modeling technique is quick, robust, does not make any assumption as to the distribution of the data, can include both categorical and continuous predictor variables, and can be used to predict both categorical and continuous type models. In this section we will be using a Random Forest modeling algorithm to predict “Land cover” types given summarized derivatives of “PCA_30” values.

From the “Statistical Modeling” menu, left click the Random Forest button ().

In the “Training Dataset” combo box, select the “Segments” feature class.

Left click the query button (just below the open button ).

Write a query in order to subset the data to the records you have populated ($LC > 0$).

Click OK.

In the Random Forest “Dependent Field”, select “LC”.

Select all the fields that start with “MEAN_” within the “Independent Fields” list box.

For “# of Trees”, type 500.

Set the “Train” ratio to 0.66 (using 66% of the data).

Set the “# Variables” of randomly selected to build each tree to 3, which is $(\sqrt{9})^6$.

Make sure the Regression check box is unchecked.

Check the “Variable Importance” check box.

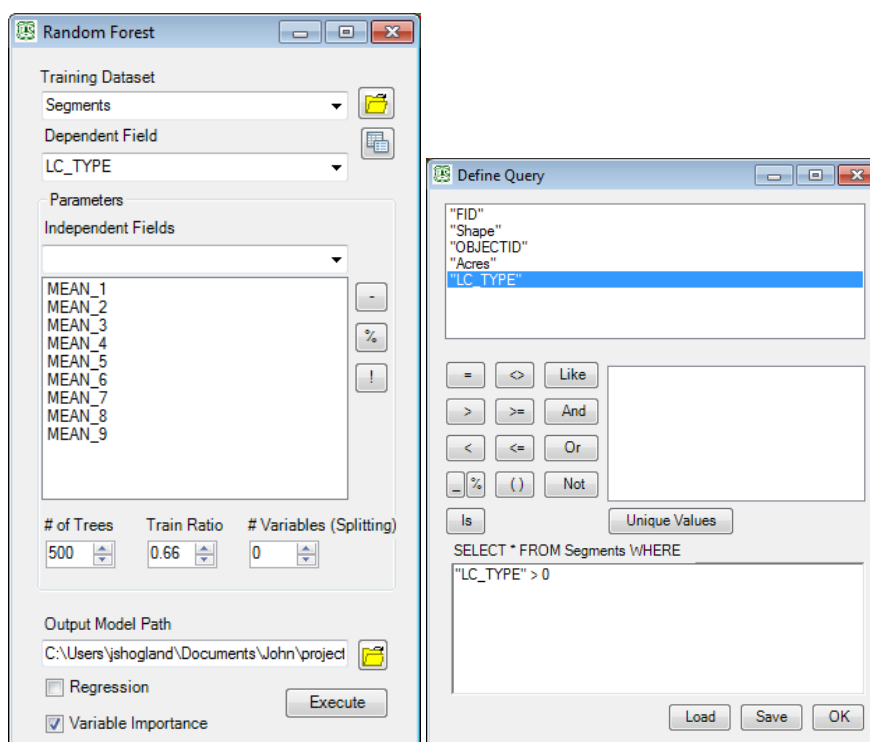
Specify a model output path to the ‘Models’ folder.

Name the model “RF_LandCover”.

Click “Save”.

⁵ Breiman L. (2001). Random Forest. Machine Learning, 45: 5-32.

⁶ Rules of thumb for proportion of data used in building the model, 0.33-0.66 depending on how noisy the data are. The noisier the data the lower the proportion. Rules of thumb for number of variables used when building the model ($\sqrt{p}$, $\frac{p}{2}$, and $p/3$).



Click “Execute”.

The Random Forest tool will ask if you want to build probability graphics.

Click “Yes”.⁷⁸

Setting these parameters and clicking the execute button will create a random forest model, report, and two graphics.

The random forest report, estimates some basic statistics related to modeling accuracy for both the data used to train the model (66%) and the “out of bag” (OOB) data (34%). These statistics can be used to evaluate the predictive capabilities of the model and contain estimates of root mean squared error (RMSE), Average error, Average relative error, Average cross entropy error, and Relative classification error for both the training data and OOB data. Depending on whether we are modeling a categorical response such as “Land cover” or we are modeling a continuous response such as average diameter, these statistics have different meanings (Table 2).

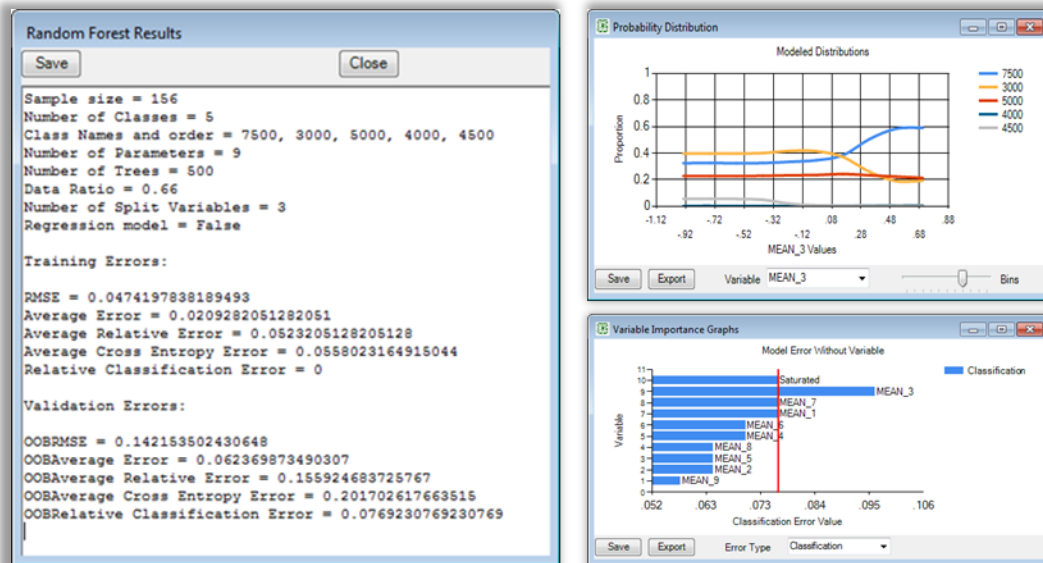
Table 2. Meaning of random forest statistics.

Categorical	RMSE	error when estimating posterior probabilities
	Average Relative Error	average relative error when estimating posterior probability of belonging to the correct class
	Average Error	average error when estimating posterior probabilities
	Average Cross Entropy Error	$\text{CrossEntropy}/(\text{NPoints} * \text{LN}(2))$
	Relative Classification Error	percent of incorrectly classified cases

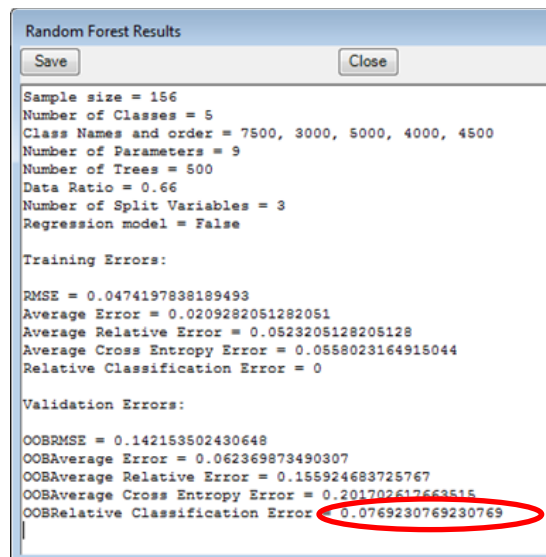
⁷ Note, to build graphics you must download and install Microsoft .Net Chart tools 3.5 sp1.

⁸ If you run into an error, check your data. DB null values and Null values will cause errors.

Continuous	RMSE	Root mean squared error of predictor value
	Average Relative Error	Relative squared error
	Average Error	Mean error
	Average Cross Entropy Error	NA
	Relative Classification Error	NA

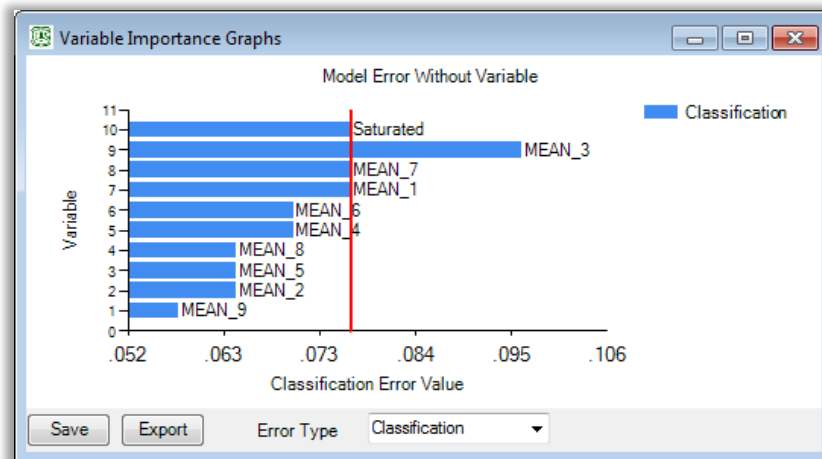


From our random forest model report we see that our model almost perfectly fits our training data. That is to say there is 0% relative classification error. We know however that it is very unlikely that our model is 100% correct and is an artifact of how random forest splits data. A better set of statistics to look at are the OOB statistics. These values are more representative of independently validated map accuracy (relative classification error of 7.6% or a map accuracy of 92.4%).

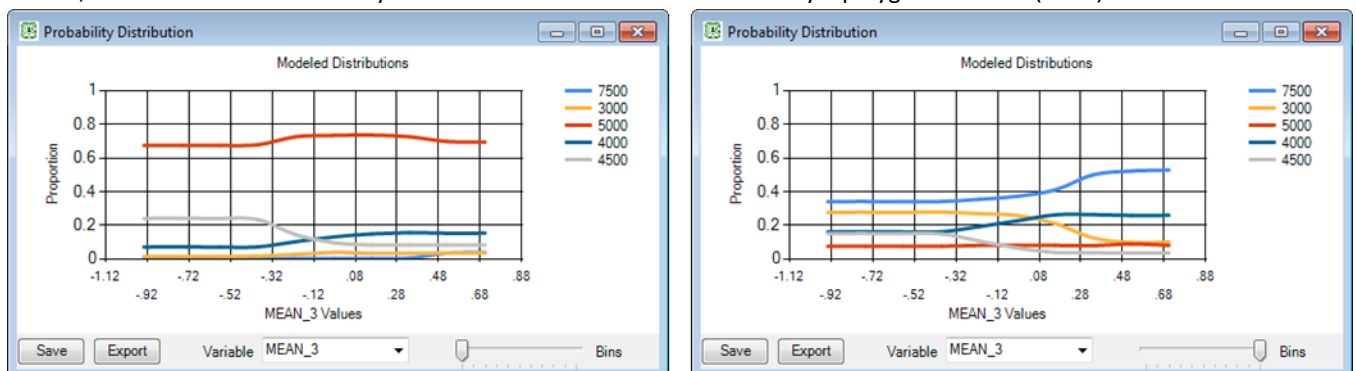


In addition to the report, the random forest tool also creates a set of graphics that can be used to visualize the importance of a given predictor variable and the functional response of the classification tree model probabilities

for each land form type. These graphics are interactive so users can look at the relationship between the response and predictor variables across a range of values. To view the importance of a given variable, users can select a given type of error and compare the increases in that chosen type of error to the error associated with the saturated model (model with all parameters). When the “Variable Importance” checkbox is checked in the random forest form, a separate classification model will be created for each predictor variables in the model. Each model removes a given predictor variable from the random forest model and records the OOB error values. Those error values are then graphically display in relationship to one another and can be used to determine the importance of each variable. In the figure below we can see that the mean value of principle component (PC) 3 has the largest classification error when removed from the random forest model.



This suggests that PC 3 is a very important variable in the model and removing it from the modeling process can increase the classification error by approximately 2.2%. Given that we now know that PC3 or “MEAN_3” has a large impact on the classification rate, let’s look at its functional relationship to each of the Land cover types. To do this we need to select the “MEAN_3” variable in from the “Variable” combo box of the “Probability Distribution” form and slide the “Bins” slider back and forth to look at the impact a given predictor variable has on a given category. In this example we see that when the “Bins” slider bar is far to the right and MEAN_3 values are small, it is more probable that the random forest models will identify the polygon as Urban (7000). As we move the slider bar to the far left, we see that it is more likely that random forest models will identify a polygon as water (5000).



The “Bins” slider bar allows users to keep all other variables used in the model at a set value based on the range of values used in the model. Low values are specified by moving the slider position to the left while higher values are set by moving the slider to the right.

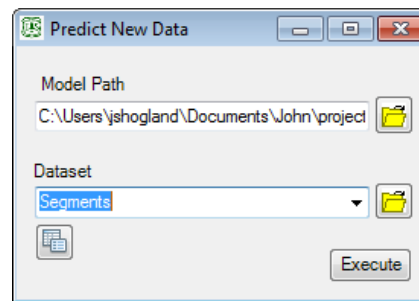
Now that we have built a random forest model let's apply it to the rest of the observations (polygons) within our "Segments" feature class. To apply our model to the Segments feature class:

Within the "Statistical Modeling" menu, click on the "Predict New Data" button (.

In the "Model Path" textbox, select the "RF_LandCover.mdl" file.

From the "Dataset" dropdown menu, select the "Segments" feature class.

Click "Execute".



Open the Segments attribute table and observe that six new fields have been added to the "Segments" feature class. Five of these new fields are named after each one of the given Land cover class codes with a prefix of "p_" and one is named after the dependent field used to build the classification with a suffix of "_mlc". The fields with a prefix "p_" quantify the proportion of Random Forest trees that classified a given record as belonging to a given class, denoted by the field name minus the prefix. The field with the suffix of "_mlc" identifies the most likely class of that record, determined from all random forest trees (the largest proportion).

SECTION 5: INTERPRETING THE CLASSIFICATION (INTERPRETATION)

Congratulations, we have now created a Land cover classification. So what does that mean and how should we interpret the outputs? The short answer is that we now have a new map that identifies the most probable "Land cover" class given the predictive variable and that on average (assuming our data are representative of the population) our Land cover map is approximately 92% accurate. But, which classes are being confused and what if our sample was not representative of the population? The first step to answer these questions begins with subjectively evaluating the classification results by visually looking at how multiple polygons were labeled.

In these next steps, we'll display the Segments layer using unique categories so we can visually compare the NAIP imagery with the Land cover classifications assigned to each polygon. This is how we'll subjectively assess the accuracy of our Land cover map.

Move the "NAIP_30" raster dataset to the top of ArcMap's "Table of Contents".

Move the "Segments" layer right below the "NAIP_30" layer.

Right click on the Segments layer.

Click on "Properties".

In the "Symbology" tab below "Show:" (on the left), click "Categories".

Select "Unique values".

In the "Value Field", select "LC_mlc".

Click "Add All Values".

Unselect "<all other values>".

Click OK.

Now, open the "Image Analysis" window (click on the ArcMap "Windows" menu, select "Image Analysis"). Highlight the "NAIP_30" raster dataset.

Left click on the "Swipe Layer" button () within the "Display" section of the "Image Analysis" window. Zoom into an area of interest.

This will put an arrow on your main screen which you can use to swipe the NAIP_30 raster dataset on and off and look at how each polygon was labeled in relationship to the NAIP_30 raster dataset.

Often, visual evaluations represent the end of a classification. However, a more formal accuracy assessment can be performed to determine the accuracy of the classification. Accuracy assessments compare known locations (Reference) of Land cover types against predicted "Land cover" types (Mapped) for a randomly chosen subset of the population. These datasets, often called validation datasets, are expensive and time consuming to acquire. Fortunately, these data have been collected and saved as a feature class called "Validation". The "Validation" dataset contains 174 randomly selected polygons from the original "Land cover" classification that have been independently evaluated. The results of the field visits are stored in the field called "Reference". The "Org_ID" field within the "Validation" feature class links to the "Objectid" field within the "Land cover" feature class and can be used to join the mapped values within the "Land cover" dataset to the "validation" dataset. To perform the accuracy assessment:

First, we'll add a new field called "Mapped" to the "validation" dataset.

Navigate to the data directory and locate the "validation.dbf" dataset.
Add the "validation" data set to the Table of Contents (you can drag/drop).
Open the "validation" attributes table.
Click the "Table Options" button at the top left of the Table field.
Click "Add Field".
Name the field "Mapped".
For "Type:", select "Long Integer".
Click "OK".

Next, we'll join the "Segments" dataset to the "validation" dataset.

Right click on the "validation" layer in the Table of Contents.
Hover the mouse over "Joins and Relates".
Click "Join."
For field 1., select "OBJECTID"
For field 2., select "Land cover".
Check "Keep all records".
Click "OK".

Next, we'll populate the "Mapped" field with the values of the "LC_mlc" field within the "Segments" feature class.

In the "validation" attribute table, right click the "Mapped" column.
Click "Field Calculator"
Double-click "Segments.LC_mlc".
It should now say "Mapped = [Segments.LC_mlc]".
Click "OK".

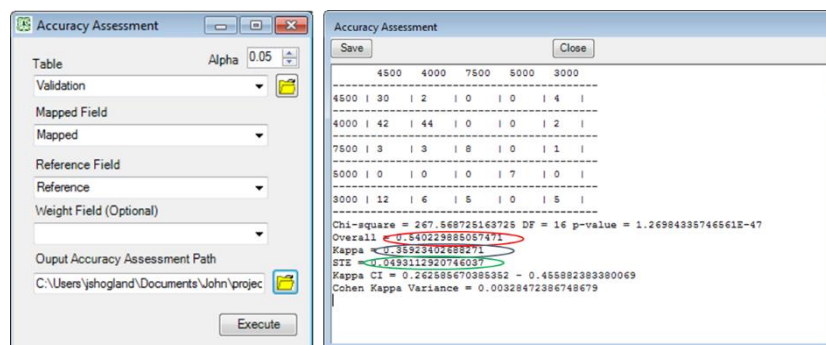
Then, we'll remove the join.

- Right click the "validation" layer.
- Hover the mouse over "Joins and Relates".
- Click "Remove Join(s)"
- Click "Remove All Joins".

Now, each record has a mapped value in the validation table and we can perform the accuracy assessment.

- Within "Statistical Modeling" menu on the RMRS toolbar, open the "Accuracy Assessment" form.
- For the "Table" combo box, select the "validation" feature class.
- For the "Mapped Field" field combo box, select "Mapped".
- For the "Reference Field" combo box, select "Reference".
- Leave the "Weight Field (Optional)" combo box blank.
- Specify an output model in the "Output Accuracy Assessment Path" (model called LC_AA.mdl).
- Click "Execute".

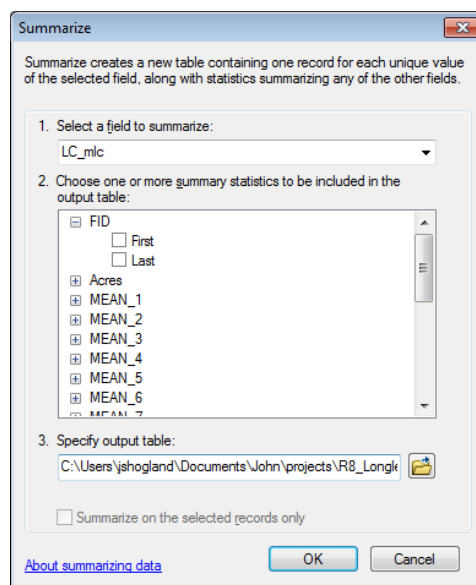
The "Accuracy Assessment" report depicts the number of correctly and incorrectly classified sampled observations (error matrix) and multiple statistics. Reference "Land cover" types are presented across the top of the error matrix and represent actual "Land cover" types that have been field verified. Mapped "Land cover" types (what the classification model predicted) are presented along the left side of the report. The counts of observations falling into each cell identify when the classification model correctly (along the diagonal) and incorrectly (off diagonal) allocated a "Land Cover" type. Below the error matrix⁹, we can see an overall accuracy of 54% (Overall, red circle) and an estimate of agreement of 0.35 (Kappa, purple circle) with a standard error of 0.049 (STE, green circle). Note, this independent map accuracy is much less than our estimate of map accuracy given our OOB relative classification error (8% -> 92% map accuracy). This is because our original sample was not a random sample and was not designed to match the population.



Now let's summarize our data by class and acreage and build a bar graph depicting the number of acres within each "Land cover" type.

⁹ In addition to the basic statistics talked about here, the report for an accuracy assessment creates the full error matrix. Reference classes are displayed across the top as columns while mapped classes are displayed along the left side as rows. To view this matrix, expand the report by left click, dragging out the corners of the form.

Open the "Segments" feature class attribute table.
 Right click on the "LC_mlc" column.
 Left click on the "Summarize..." button.
 Expand the "ACRES" field.
 Check "Sum".
 Navigate to data directory.
 Name the output table "AcreageByLcType.dbf".
 Click "OK".
 It will ask if you want to add the result table in the map.
 Click "Yes".



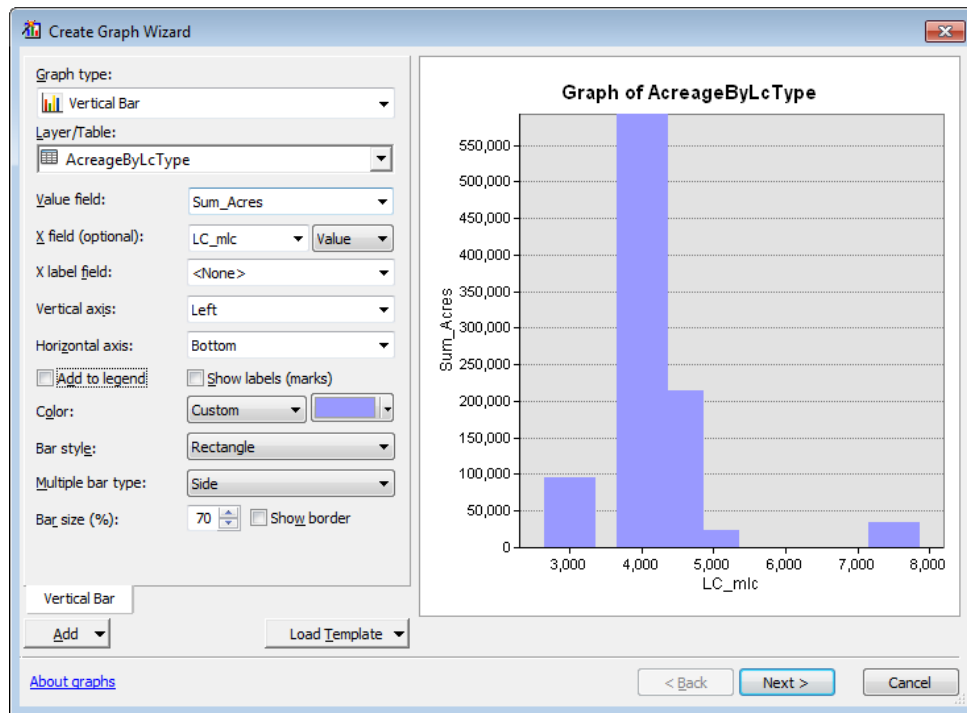
Left click on the ArcMap menu "View".
 Hover your mouse over "Graphs".
 Left click "Create Graph".

These actions will bring up ArcMap's "Create Graph Wizard".

Select a "Vertical Bar" graph type.
 For "Layer/Table:" select "AcreageByLfType".
 For the "Value field:", select "Sum_ACRES".
 For the "X field (optional):", select "LC_mlc".
 Uncheck the "Add to legend" checkbox.
 Click "Next".
 Click "Finish".

These actions will create a graphic window which can be added to a layout.

Right clicking within the graphic.
Click “Add to Layout”.



Let’s complete a layout displaying the spatial location of our Land cover types. You can add various components to your layout display. If you’re already familiar with this process, go for it! If you’re not certain how to complete your layout, we’ll go through a few key steps.

Let’s remove the polygon outlines from our Segments layer:

- Open the “Layer Properties” box for the “Segments” layer.
- Navigate to the “Symbology” tab.
- Under “Show:” select “Categories”.
- Right click on the Segments classifications.
- Left click “Properties for All Symbols”.
- Under “Outline Color”, select “No Outline”.

To insert a title:

- Left click on the ArcMap “Insert” menu.
- Left click “Title”.
- Type in “Land cover Map” as the title.
- Click “OK”.

To insert a scale:

- Left click on the ArcMap “Insert” menu.
- Left click on “Scale Bar”.
- Select a scale of your choosing.

Click "OK".

To insert a compass rose:

Left click on the ArcMap "Insert" menu.

Left click "North Arrow".

Select a compass rose of your choosing.

Click "OK".

To insert a legend:

Turn off all layers in the Table of Contents except "Segments".

Left click on the ArcMap "Insert" menu.

Left click on "Legend".

Double check that "Segments" is listed in "Legend Items".

(If you need to remove a "Legend Item": highlight it and click the left arrow button located in the middle.)

Click "Next" four times.

Click "Finish".

The legend will be added to the layout. You'll notice, however, that it is still displaying the Land cover classification codes (e.g., 3000) rather than the names of the classes (e.g., "Grass"). To change these labels:

Open the "Layer Properties" box for the "Segments" layer.

Navigate to the "Symbology" tab.

Under the heading "Label", click on a Land cover classification code (e.g., 3000).

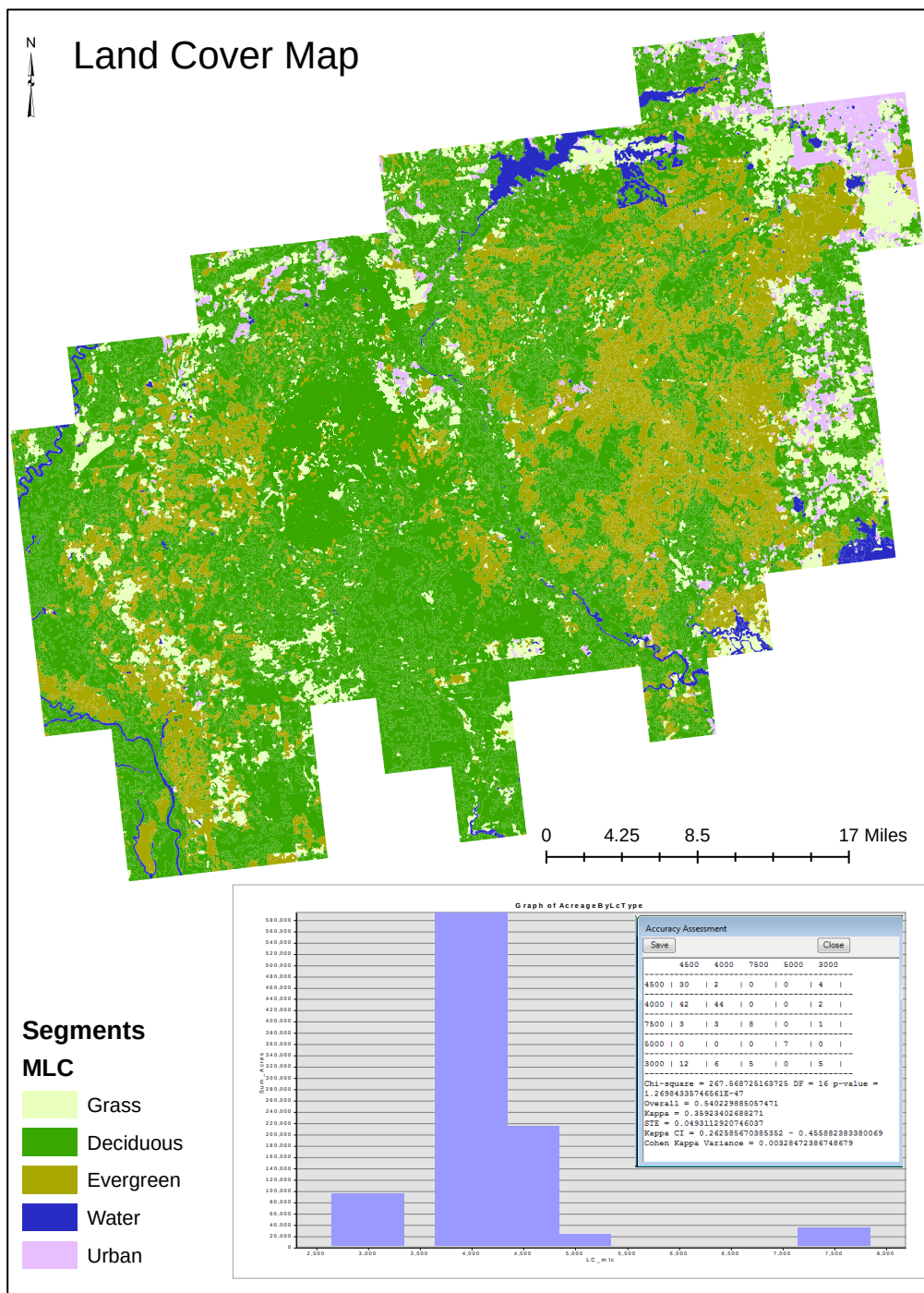
Change the classification code to the name of the appropriate class (e.g., change 3000 to "Grass")

Repeat for all Land cover classes.

Click "OK".

Now we have a layout of our Land cover Map. We kept our layout elements simple. However, you can customize your layout, move these elements around, alter their properties and display details (e.g., change the symbol colors), add extra elements (e.g., a graphic of your accuracy assessment), and tailor your layout to the findings you intend to report.

See below for an example of a Land cover map layout.



When finished, click “File”, then “Export” to export the newly created map to an image or pdf file format and report your findings.

This section completes the “Classification” tutorial. While we have described how to create a random forest classification using imagery and visually interpretation, there are many more data mining and statistical tools that can be used within RMRS Raster Utility toolbar. Each tool works similar to the tools used in this tutorial and provides a framework to explore, describe, and test hypothesis. Happy modeling!