Pontus Olofsson, Christopher E. Holden, Eric L. Bullock
Boston Education in Earth Observation Data Analysis/
Department of Earth & Environment, Boston University

S5. Methods: Object Based Analysis

S3.1 Create a spatial subset to work with (optional)

Since image segmentation and object-based image analysis (OBIA) is computationally expensive, it is best to work on a subset to determine an appropriate workflow. This makes it quicker and easier to see the effect of changing segmentation algorithms, algorithm parameters, training data, and classification parameters. Since every location is different and every project demands unique outputs, no single set of parameters can be applied to every project. Therefore it is a dynamic process. The subset used should be big enough to include all land cover types that should be mapped in the end product, but small enough to drastically reduce computation time. It is also recommended, if possible, to keep the relative proportions of classes similar to the full image for classification purposes. Not that this is only necessary for large, high-resolution images.

1. Open the image in QGIS and navigate to an area you wish to subset.
2. Create a subset by going to *Raster -> Extraction -> Clipper.*
3. Select the raster image for the *Input file* and specify and output name. Indicate -9999 as the *No data value* unless you know the input no data is different.
4. Select *Extent* for the *Clipping mode* (it should be already highlighted).
5. Click and hold down on the map on the corner of where you wish the subset to be. Drag your mouse while continuing to be holding down and a red box should appear. This box indicates the area for your subset, so make it the area you desire.
6. Click *OK* and a subset of your original image should load to your *Layers*.

S3.2 Segment the image

There are various methods of image segmentation that can be used for object-based analysis. The method that we will be using, due to its' availability in the Orfeo Toolbox is called Mean-Shift Segmentation. Mean-shift segmentation is a non-parametric clustering technique that is widely used in image processing. The algorithm assumes that the spectral values in the image are sampled from an underlying probability density function. The dense clusters of data are therefore assumed to correspond with the modes of the underlying density function. The mean spectral values are computed inside a moving window for each image. The window is then moved to the mean and the process is iterated until the means converge. The algorithm makes no assumption about the number of modes or size of output clusters.

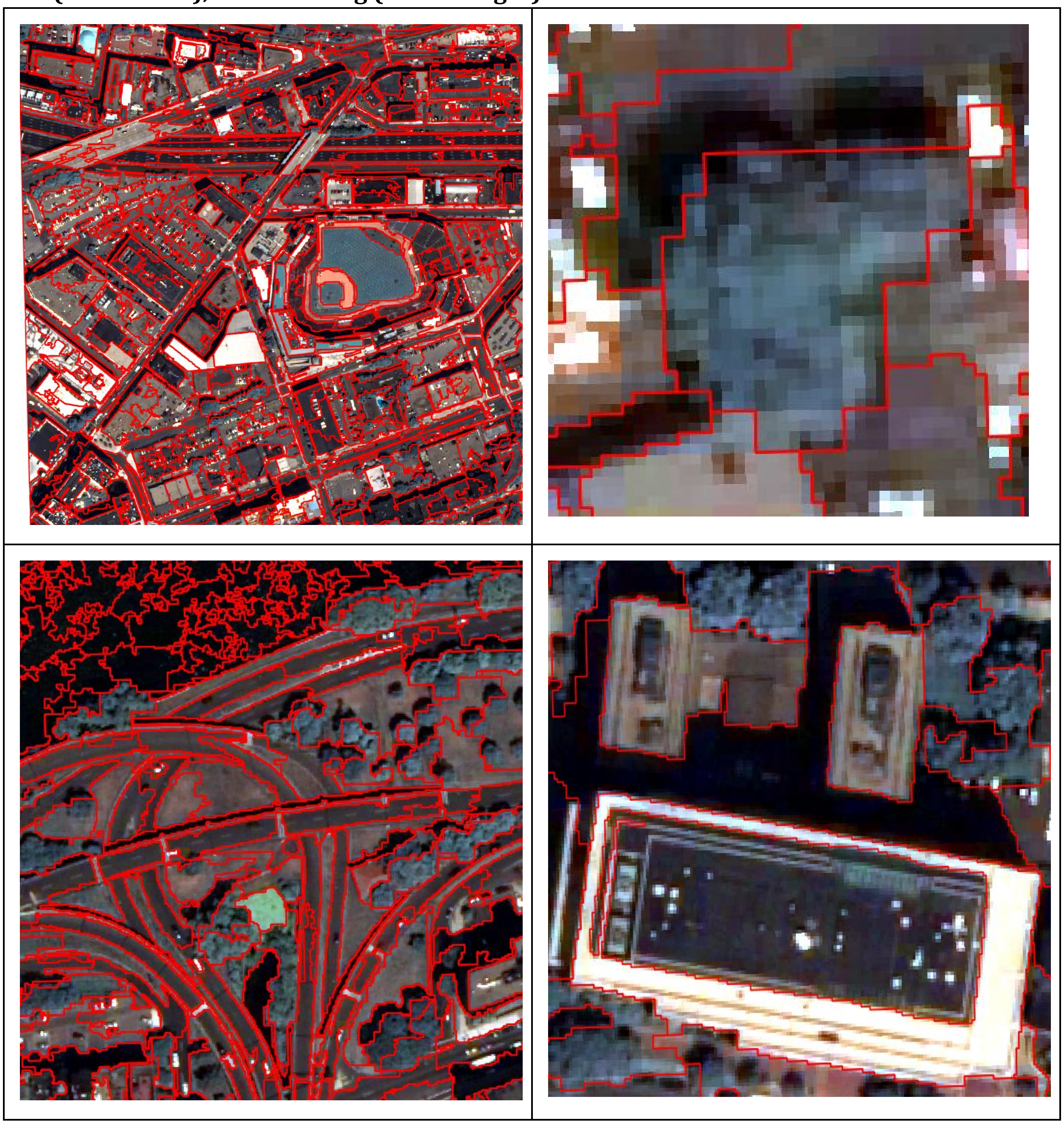
1. In the *Processing Toolbox* (Processing -> Toolbox), go to *Orfeo Toolbox -> Image Filtering ->*

Mean Shift filtering.

2. Select your image for the *Input Image.*
3. You might have to tweak the parameters a bit to get the results you are looking for. Here are some suggestions:
    a. *Spatial radius*: the size of a window (in pixels) being considered in mean calculation; to start leave this at 5.
    b. *Range radius:* specifies how close, in terms of euclidian distance, neighboring pixels need to be to be grouped together; to start, try putting it at 30.
    c. *Mode convergence threshold:* Neighboring pixels whose multi-spectral distance lies below this threshold with converge into one mode. Leave this at .1.
    d. *Maximum number of iterations:* Algorithm will stop if convergence has not been reached at this amount of iterations. This can impact computation time. If you want to speed up the process at the expense of segment quality, decrease this value. To start, keep this at 100.
    e. *Range radius coefficient.* This must be set to 1 or the segmentation will not work.
4. Turn *Mode search* to off.
5. Save the *Filtered Output* to "Filtered.tif".
6. Save the *Spatial Image* to "Image.tif".
7. The reason multiple files are created is because the image is broken up into tiles during segmentation. If the segment tiles were to be put back together, there would be obvious seams in the segments where the tile borders were. The images created here are used later to retain the spatial homogeneity of the image and prevent these tile seams. Keep everything else the same, and click *Run.* Depending on the size of your input image, this could take a little while.
8. Move on to part 2 of the segmentation process by going to *Orfeo Toolbox -> Segmentation -> Exact large-Scale Mean-Shift segmentation, step 2*.
9. For the *Filtered image* and *Spatial image* select the 'Filtered.tif' and 'Spatial.tif' images respectively.
10. For the various parameters:
    a. *Spatial radius* and *Range radius:* Select the values you selected in step 1.
    b. *Minimum region size:* Any segment under this size in pixels will be deleted after this step. Keep this at 0.
    c. *Size of tiles in pixels:* To save on memory, the image is broken up into tiles. If the program is crashing, reduce these values. To start, however, keep them at 500.
11. Save the *Output Image* as Output.tif and click *Run.*
12. Optionally, you can limit the region size for your segments and join small segments with their appropriate neighbor. Do this by going to *Orfeo Toolbox -> Segmentation -> Exact Large-Scale Mean-Shift segmentation, step 3 (optional).* The *Input image* is your original image, the *Segmented image* being the 'Output.tif' file created in the previous step, and the *Minimum Region Size* being the minimum size you wish the segments to be. Save the output as 'Output_Step3.tif'. Click *Run*.
13. To finalize the segmentation and output a vector file to work with, go to *Orfeo Toolbox -> Segmentation -> Exact Large-Scale Mean-Shift segmentation, step 4*. For *Input Image*, use the

original image you are segmenting. For *Segmented Image,* use either the output from the 3rd segmentation step (#12 in this section), or the second step (#8), depending on whether you did the optional region merging step. Save the *Output GIS vector file* to a shapefile such as 'Segmentation.shp' and click *Run.*

Figure 1. Examples from Boston. A zoomed-out subset (top left), tree (top right), freeway exit (bottom left), and building (bottom right)



S3.3 Merge segments by hand (optional)

Depending on how you set your parameters, you may have more segments than you have features. This will happen especially when you set your minimum object size to a small number to try to retain smaller features such as trees. One optional step after segmentation is to manually merge adjoining segments when you feel it unnecessary to have multiple objects for the same feature.

1. Highlight your segment polygon and set it to editing mode (right click -> *Toggle Editing)*.
2. Highlight the different segments you wish to merge while holding the 'Control' key on the keyboard. This allows multiple objects to be highlighted at once.
3. Go to *Edit -> Merge Selected Features.*
4. A screen should appear asking which Digital Number values to obtain for the merged feature. In our case this is irrelevant, so click *OK.*
5. Repeat this step for any segments you wish to merge.
6. Toggle editing once again and click *Save.*

Figure 2. A road being merged from two segments into one.

Before	After
	

S3.4 Spectral statistics

Automatically, the shapefile output from the segmentation contains segment spectral statistics in its' attribute table. For each segment and for each spectral band in the original image, there is an attribute field for mean and variance of all pixels within the segments.

1. If it is not already added, add the shapefile to QGIS by going to *Layer -> Add Layer -> Add Vector Layer…* and selecting your shapefile.
2. In the *Layers* panel, right click on the shapefile and select *Open Attribute Table.* You will see field names such as meanB0, meanB1...varB0, varB1, etc. Since the algorithm behind this is written in Python, and Python is indexed on 0 instead of 1, B0 actually represents

Band 1. Therefore, B1 represents Band 2, B2 represents Band 3, and so on.

3. Close the Attribute table.
4. Normally, these spectral statistics are enough to represent the majority of the spectral information contained in the segments. If this is your first time going through this, **skip to the next section**. If you are not happy with your results are are exploring more variables to represent the segments, you can do zonal statistics to calculate more statistics besides mean and variance.
5. There are more zonal statistics you can calculate in QGIS, but you have to do it by hand and individually for each band. The options that are available are: count, sum, mean, median, standard deviation, minimum, maximum, range, minority, majority, and variety.
6. If one of these statistics would be especially useful as a statistic, you can use the 'Zonal Statistic' plugin to calculate them.
    a. First, activate the plugin by going to *Plugins -> Manage and Install Plugins* and checking the white box next to 'Zonal Statistics Plugin'. Click *Close.*
    b. Open the plugin by going to *Raster -> Zonal Statistics -> Zonal Statistics.*
    c. For *Raster layer* select your input image.
    d. For *Band* select the spectral band you want to do the zonal statistics on.
    e. For *Polygon layer…* select the segmentation vector.
    f. For *Output Column Prefix* put something to indicate the statistic being calculate. **Note:** For the classification script to work, this prefix cannot contain any symbols. An example for calculating statistics on Band 1 would be 'newB1'.
    g. Select which statistics the calculate, a new attribute field will be created for each one.
    h. Click *Okay.* Repeat this for as many bands as you wish.

S3.5 Calculate segment geometric statistics (optional)

One advantage when working with objects instead of pixels is that you can take advantage of the geometric properties of the objects in addition to their spectral signatures. While this is not always necessary, it may help in distinguishing between similar land covers. An example would be sidewalks and buildings. Sidewalks are often long and skinny, while buildings tend to be more cubicle or round in nature. There are many geometric calculations you can do, but here we will stick to 3 to keep it simple: area, perimeter, and roundness.

1. Area
    a. Open the attribute table for the segmentation vector by right clicking the layer in the *Layers* toolbox and going to *Open Attribute Table.*
    b. Toggle the editing mode by clicking the XX button. You can now make changes to the vector.
    c. Open the *Field calculator* by clicking the XX button.
    d. Under *Output field name* type 'Area'.
    e. Under *Function List* navigate to 'Geometry' and double click '$area'. The *Field*

calculator should now look like this: XX

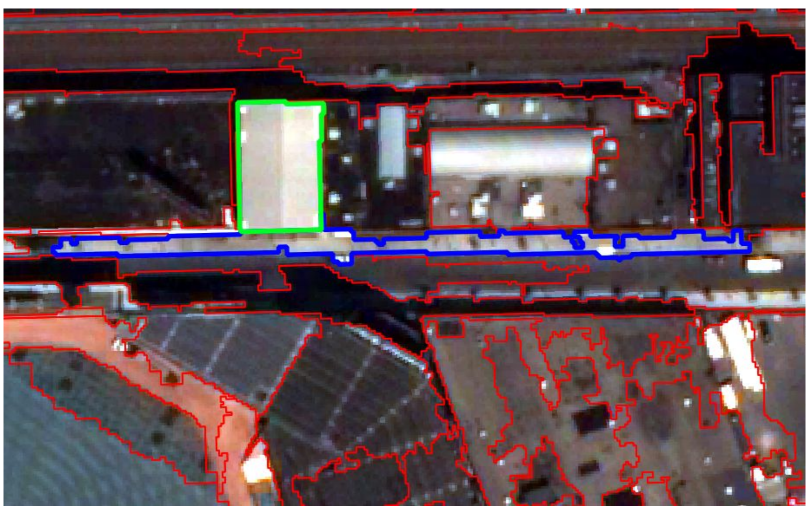
f. Click *OK.* A new attribute will be added to the attribute table for the file.

   i. Note: If your using a non-projection reference system, it must be converted into a projected one. If you are receiving unusual values for the area or an error message, try changing the CRS. Do this by going to *Vector -> Data Management Tools -> Define Current Projection*. Pick a projected CRS such as WGS84 UTM##.

2. Perimeter
   a. Return the the *Field calculator.*
   b. Under *Output field name* type 'Perimeter'.
   c. Under the *Function List* navigate to 'Geometry' and double click '$perimeter'. Click *OK.*
3. Roundness
   1. In the *Field calculator* type the following equation:

      (4 * $pi * $area)/($perimeter^2) * 1000

   2. **Note**: The '$pi' field can be found in the 'Math' category.
   3. Name the field 'Perimeter' and click *OK.*



	Sidewalk	Building
Area	640	588
Perimeter	378	200
Roundness	56	184

Figure 3. A building is highlighted in green, and an adjacent sidewalk in blue. While spectrally they are similar, their shapes (especially perimeter and roundness) differ substantially.

S3.7 Collect training data

Any sort of automated image classification involves using labeled training data as inputs to some sort of classifier. This classifier can then be used to label the rest of the image. This is the same process for OBIA as it is with pixel-based analysis, only the training samples are objects instead of pixels.

1. It will be easier to find segments to use as training inputs if the vector is overlayed on the original image. In the *Layers Panel* make sure only the segmentation shapefile and original image are checked, and move the segmentation vector on top of the image so that it is visible.
2. Check the symbology so that the fill is transparent but you can still see the segment outlines.
    a. Right click on the segmentation vector and go to *Properties.*
    b. Go to *Style* and highlight *Simple fill*, which can be found under *Fill,* by clicking on it.
    c. Select the down arrow next to *Fill* under *Symbol layer type* and select 'Transparent Fill'.
    d. Click *Okay.*
3. Add a 'Class' field to the attribute table of the segmentation vector.
    a. Right click on the layer and go to *Attribute Table.*
    b. Click the pencil icon (XX) to begin editing.
    c. Select the *New Attribute* option (XX) and add a field named 'Class'. Click *Ok* to add the field.
4. In the bottom-left corner of the *Attribute table* change the view option to 'Show Selected Features'. Return to the image viewer, but do not exit the *Attribute table.*
5. Start with one class, and find representative examples of segments in that class. Turn feature selection on by selecting the 'Select Features' (XX) tool.
6. Click on the segment while holding down the Command (or Windows) key on the keyword. While holding down the key, select more segments. You should notice that because you are holding down on the keyboard, you can select multiple features at the same time.
7. Once you have an adequate number of samples in this class, return to the *Attribute table*. It is hard to say how many features should be used. This depends on size of your image and amount of segments. In general the samples should be representative of the class in all locations in the image, and the amount should be proportional to the amount that land cover occurs in the study area. For example, if you have more forest area than water, you should collect more forest samples than water.
8. In the *Attribute table* open the *Field calculator (XX)*.
9. Select 'Update existing field' and select the 'Class' field. Make sure 'Only update X selected features' is checked.
10. In the text input box, assign an integer value to represent the class (for example, forest could be 1, water 2, non-forest 3, etc.). Click *Ok.*
11. The 'Class' field should not contain the value of the class it was assigned to for the selected segments. You can add more samples individually by selecting them on the map, and double clicking on their 'Class' attribute in the *Attribute table* like you were writing in a spreadsheet. There you can simply type in the class.
12. Repeat this for all of your classes.

S3.8 Compute Feature Statistics XML File

For the classification process, and file must be created that contains the means and variances of the input features you will use for training and classifying.

1. Go to *Processing Toolbox -> Orfeo Toolbox -> Segmentation -> ComputeOGRLayersFeatureStatistics*
2. For *Name of the input shapefile* select the segmentation shapefile.
3. For *XML file…* select a file name to save as the output. This file must have a .xml extension. An example would be 'Features.xml'.
4. For *List of features to consider for statistics*, add all of the layer features you wish to use for classification. They must be spelt exactly as they are in the attribute table, contained in quotations, and separated by a space. The entire field must be contained in single quotations. If you wanted to use the mean and variance of bands 1-7, area, perimeter, and roundness, you would write:

'"meanB0" "meanB1" "meanB2" "meanB3" "meanB4" "meanB5" "meanB6" "varB0" "varB1" "varB2" "varB3" "varB4" "varB5" "varB6" "Area" "Perimeter" "Roundness"'

5. **Note:** if you are using different fields than the ones used above, copy and paste it into a word document to make it easier to re-use later.
6. Click *Run.*

S3.9 Train the Classifier

The classifier implemented in the object classification tool in Orfeo is a Support Vector Machine. First, you need to use the training data you created to train a classifier.

1. Go to *Processing Toolbox -> Orfeo Toolbox -> Segmentation -> TrainOGRLayersClassifier*
2. For *Name of the input shapefile* select your segmentation shapefile.
3. For *XML file…* select the feature statistics XML created in the previous step .
4. For *List of features to consider for statistics,* use the same features and format used in the previous step.
5. For *Field containing the class id for supervision* put the name of the Field that holds the training class labels (ie. Class).
6. Save the output model to a file with a .svm extension. An example would be 'model.svm'

S3.10 Classify the Object Shapefile

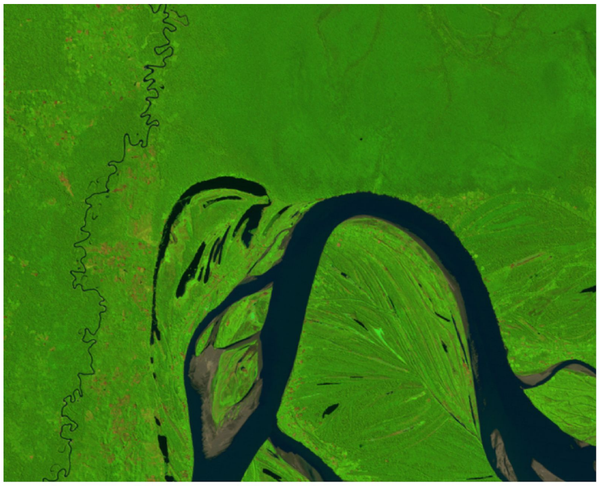
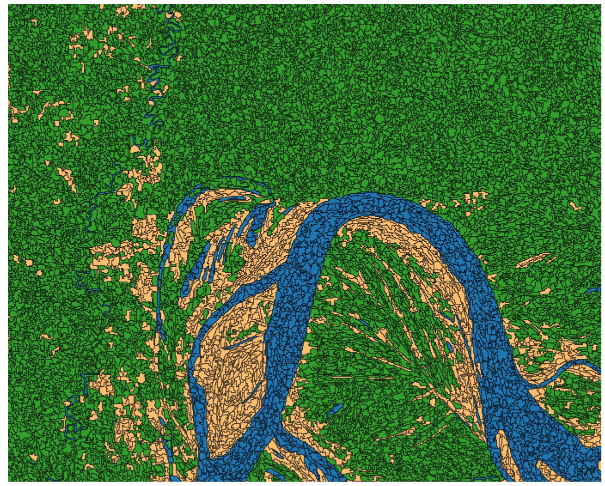
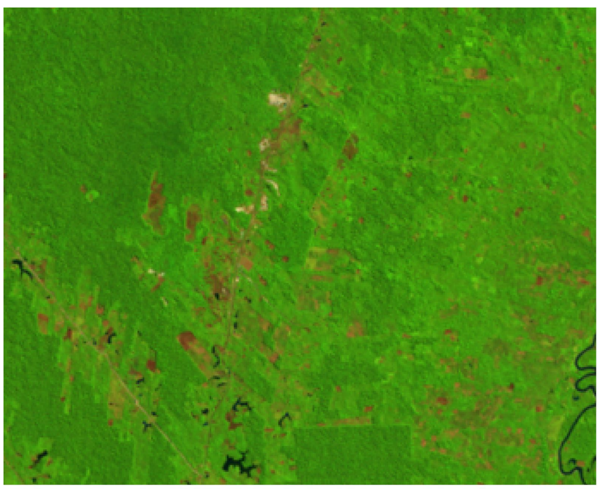
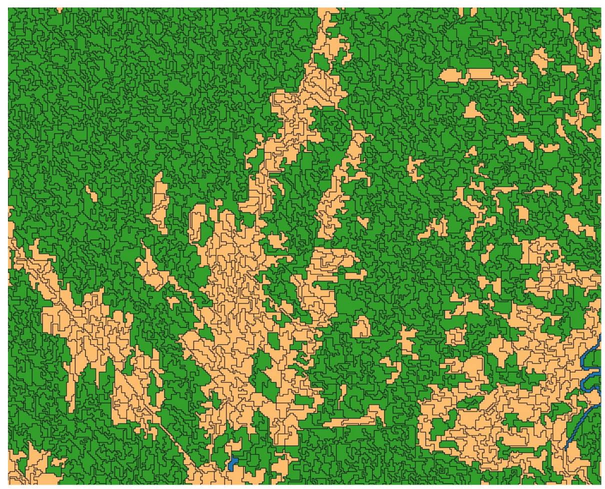
The classifier used in the previous step can then be used to classify the segments that were not labeled in the training data collection.

1. Add a ‘predicted’ field to the attribute table of the segmentation vector.
    a. Right click on the layer and go to *Attribute Table.*
    b. Click the pencil icon (XX) to begin editing.
    c. Select the *New Attribute* option (XX) and add a field named ‘predicted’. Click *Ok* to add the field.
2. Go to *Processing Toolbox -> Orfeo Toolbox -> Segmentation -> OGRLayerClassifier.*
3. For *Name of input shapefile* select your segmentation shapefile.
4. For *XML file…* select the feature statistic XML created in **S3.8.**
5. For *Features* put the feature list used in the same format and order as in **S3.9** and **S3.10**.
6. For *Field containing the predicted class* put the name of the field you just created in the attribute table (‘predicted’).
7. For *Input model filename* select the SVM classifier created in **S3.9**. **NOTE:** It will ask you if you want to overwrite the file. This is a bug, you are not actually creating a new model in this step. Just click yes and then *Run.*

S3.11 Visualize the Classification

Now that the segments are classified, you can visualize the data to represent a map.

1. Right click on the segmentation shapefile in the *Layers Panel* and go to *Properties.*
2. Go to *Style* and select ‘Categorized’ in the dropdown list at the top of the panel.
3. Select the ‘predicted’ column.
4. Click *Classify* to assign colors to each of the values. If you want to change the colors representing each class, double-click on the colored box in the ‘Symbol’ column. From here you can change the color of the segment fill, pattern, transparency, and outline color.
5. Click *OK.*

Landsat 5-4-3 composite	Resulting object-based classification
	
	

S3.11 Manual Clean-Up

The final step in the classification is to fix any segments that are visually misclassified by hand.

1. Open the *Attribute table* for the shapefile.
2. Start editing by selecting the pencil (XX) tool.
3. Manually change the class of of the ‘predicted’ column for misclassified segments.
4. Repeat until satisfied.