Project Synopsis and Purpose:

Creating a hybrid or a supervised classification in Erdas allows for a highly accurate result (or as much time as the analyst can spend on it). I wanted to investigate another approach to this – the unsupervised classification. How does a more automated approach affect the final image output? What occurs when different specifications are used for the unsupervised classification?

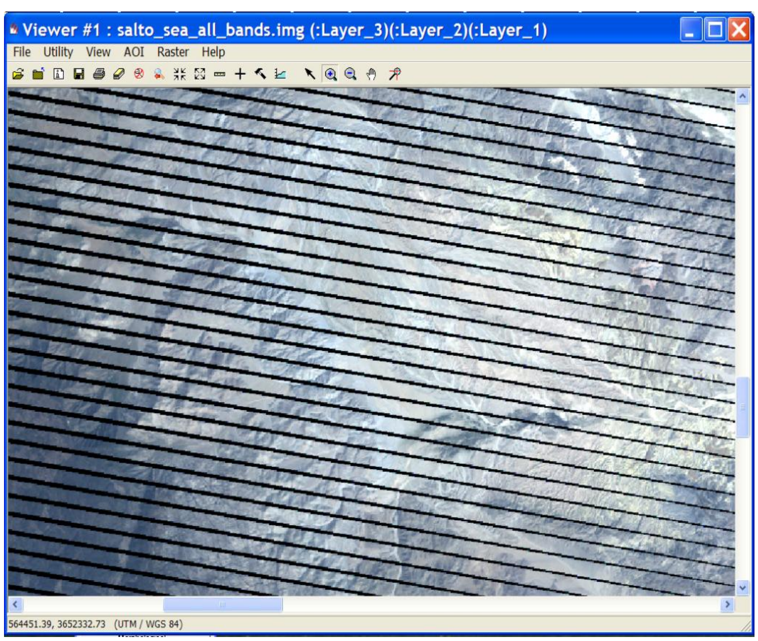
I used a Landsat image of the Salton Sea, and broke my project into two parts. The first part involved running a series of unsupervised classifications, starting with six classes and ending with thirty. Leaving all other options during the process constant, I was able to compare how the number of classes affected image output resulting from an unsupervised classification. The second part involved a simple land cover classification based on a CalFire overlay. Ten images were produced from a series of unsupervised classifications. Each image assigned an increasing number of spectral classes to each of the five classes in the land cover scheme I had created.

Data Acquisition:

The first task was to acquire an image. I wanted to use something other than the Sacramento data we had used during labs, to provide an extra challenge to my project. I chose an image of the Salton Sea; I spent about eight months in Imperial County for a conservation internship, so I'm fairly familiar with the area. A bit of research directed me to the excellent glovis.usgs.gov site.

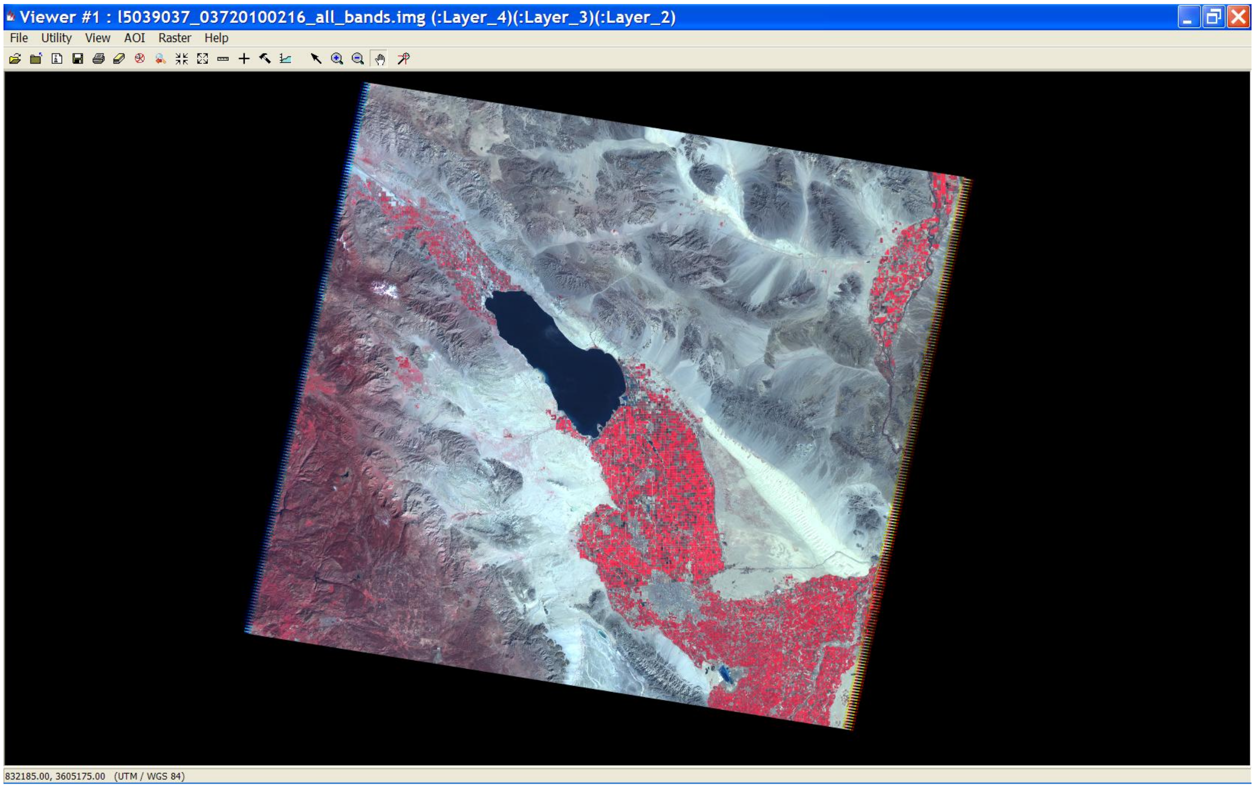
Originally I wanted to use an image from Landsat-7, but ran into some unexpected problems. The striping artifacts on the image made it useless for classification. There seemed to be ways to correct this, but I'm not familiar enough with Erdas, nor did I have enough time. Instead I chose the same scene from Landsat-5. It was very easy to find a recent, cloud-free scene of my study area. From personal experience, Imperial County rarely has cloudy days.

Striping in Landsat-7 image



After acquiring the zipped image file, I first had to find a program to unzip it (7-zip). It was interesting to find that each band was a separate file. Some quick research showed me how to prepare the bands for use in Erdas. Each band exists as a geotiff, which were imported into Erdas and converted to .img files. A layerstack was then created with the .img files, resulting in a single layer containing all seven bands. I decided to include all bands, since I wasn't sure which direction I would take my project.

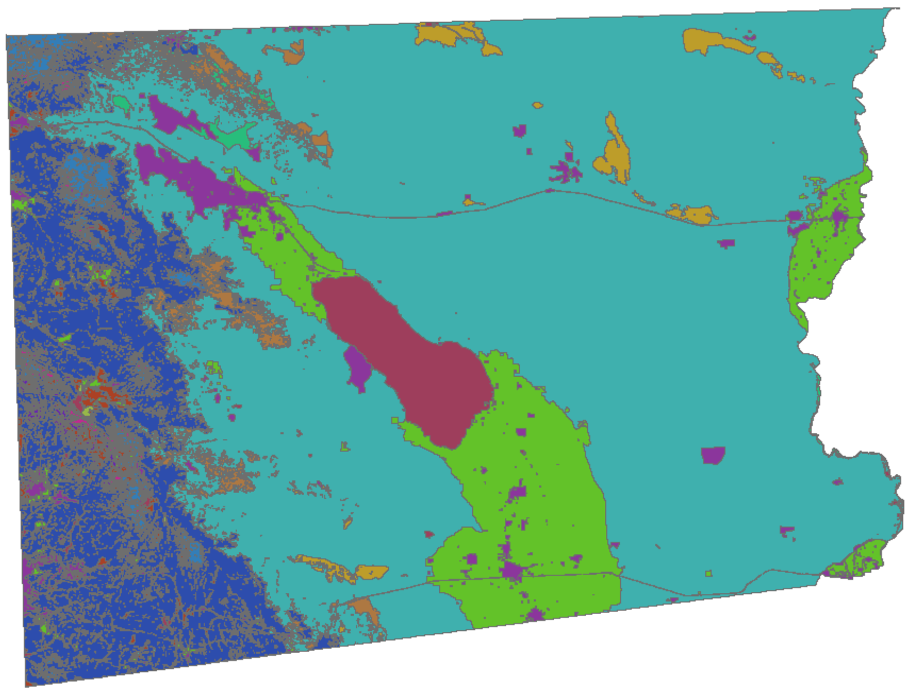
Landsat-5 scene after preparation in Erdas



The above image has a resolution of 28.5 m, and covers an area of 183 x 170 kilometers. It was acquired on February 16, 2010 by Landsat-5.

Even though I'm familiar with the area in the Landsat image, I wanted some ancillary data to gauge my results against. CalFire has a great deal of GIS land cover data available. I downloaded their Multisource Land Cover Data, a 100 m resolution grid overlay covering California. In ArcMap, I set it to display 13 major land cover classes, and clipped the grid to the outline of my study image. This was done by creating an empty shapefile, and using the Editor tool to outline the Landsat image – the resulting shapefile was used to clip the MLCD grid, and the result was exported as a shape file. The roundabout way was necessary after finding that a grid file can't be clipped by an .img file.

MLCD , clipped and symbolizing 13 land cover classes

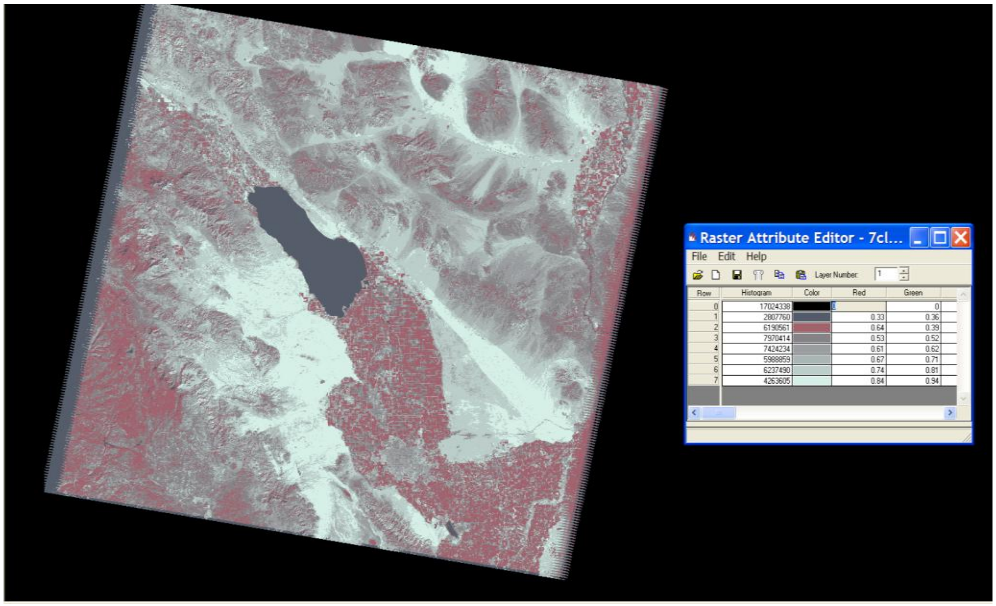


Part 1: Unsupervised Classification:

After examining the MLDC overlay and the Landsat image, I chose 6 classes as a starting point for my unsupervised classification. I needed some spectral diversity for analysis, so anything under 6 classes seemed futile. After running a few unsupervised classifications, I chose 30 classes as a stopping point. It was a good balance between proving a pattern, and having too much data.

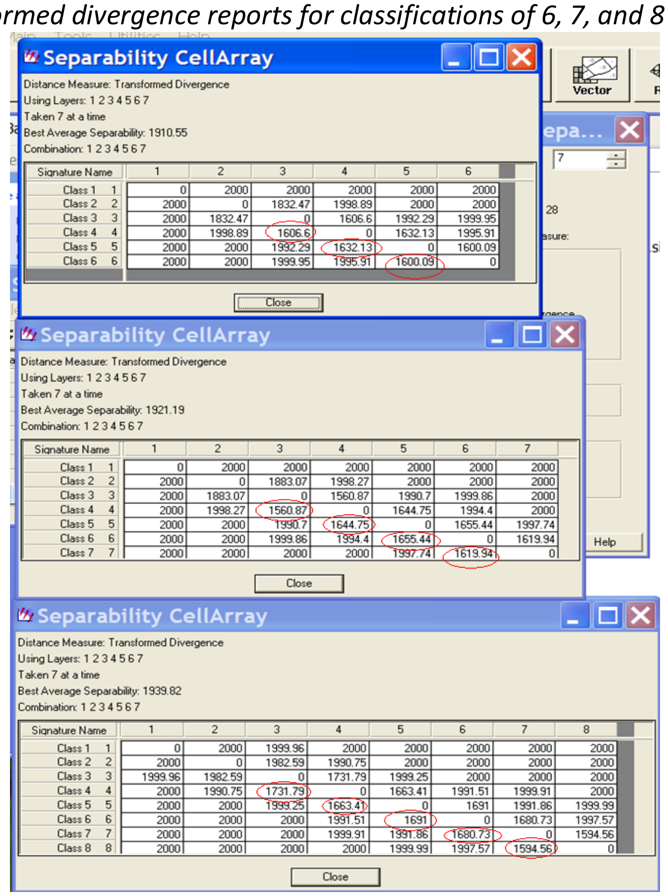
To act as controls, I maintained all defaults when setting up the unsupervised classifications. Every classification had 15 iterations, was set to maximum likelihood classifier, and was displayed as true color. Processing time was very fast. In total there were fourteen unsupervised classifications run. Transformed divergence reports and dendrograms were collected for every classification. A few signature statistics and histograms were also collected.

7 classes, unsupervised classification



After running the classifications and collecting data on them, I looked for any trends.

Transformed divergence reports for classifications of 6, 7, and 8 classes



Separability CellArray

Distance Measure: Transformed Divergence
Using Layers: 1 2 3 4 5 6 7
Taken 7 at a time
Best Average Separability: 1910.55
Combination: 1 2 3 4 5 6 7

Signature Name		1	2	3	4	5	6
Class 1	1	0	2000	2000	2000	2000	2000
Class 2	2	2000	0	1832.47	1998.89	2000	2000
Class 3	3	2000	1832.47	0	1606.6	1992.29	1999.95
Class 4	4	2000	1998.89	1606.6	0	1632.13	1995.91
Class 5	5	2000	2000	1992.29	1632.13	0	1600.09
Class 6	6	2000	2000	1999.95	1995.91	1600.09	0

Close

Separability CellArray

Distance Measure: Transformed Divergence
Using Layers: 1 2 3 4 5 6 7
Taken 7 at a time
Best Average Separability: 1921.19
Combination: 1 2 3 4 5 6 7

Signature Name		1	2	3	4	5	6	7
Class 1	1	0	2000	2000	2000	2000	2000	2000
Class 2	2	2000	0	1883.07	1998.27	2000	2000	2000
Class 3	3	2000	1883.07	0	1560.87	1990.7	1999.86	2000
Class 4	4	2000	1998.27	1560.87	0	1644.75	1994.4	2000
Class 5	5	2000	2000	1990.7	1644.75	0	1655.44	1997.74
Class 6	6	2000	2000	1999.86	1994.4	1655.44	0	1619.94
Class 7	7	2000	2000	2000	2000	1997.74	1619.94	0

Close

Separability CellArray

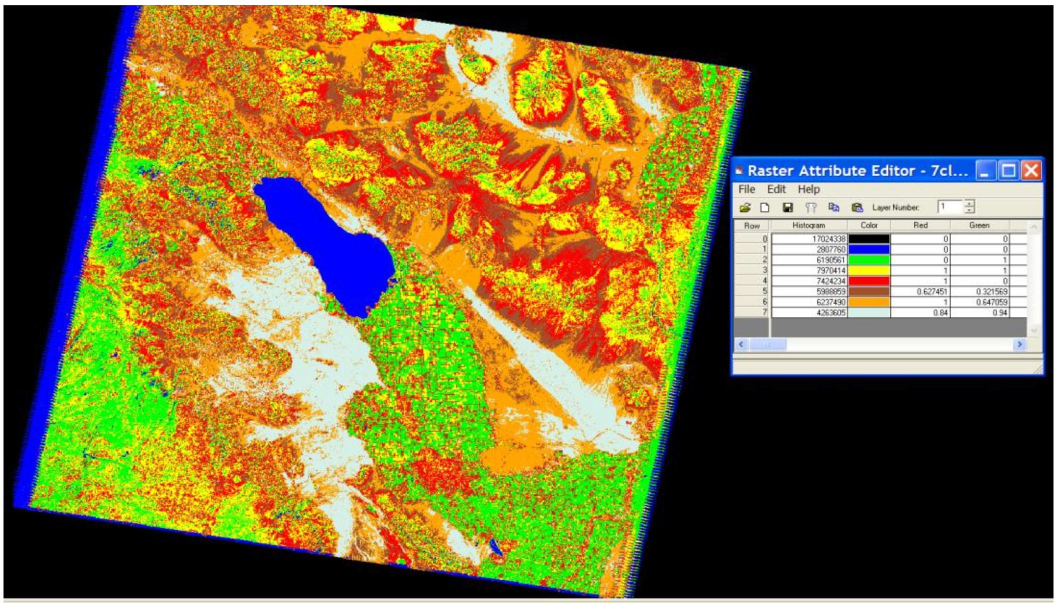
Distance Measure: Transformed Divergence
Using Layers: 1 2 3 4 5 6 7
Taken 7 at a time
Best Average Separability: 1939.82
Combination: 1 2 3 4 5 6 7

Signature Name		1	2	3	4	5	6	7	8
Class 1	1	0	2000	1999.96	2000	2000	2000	2000	2000
Class 2	2	2000	0	1982.59	1990.75	2000	2000	2000	2000
Class 3	3	1999.96	1982.59	0	1731.79	1999.25	2000	2000	2000
Class 4	4	2000	1990.75	1731.79	0	1663.41	1991.51	1999.91	2000
Class 5	5	2000	2000	1999.25	1663.41	0	1691	1991.86	1999.99
Class 6	6	2000	2000	2000	1991.51	1691	0	1680.73	1997.57
Class 7	7	2000	2000	2000	1999.91	1991.86	1680.73	0	1594.56
Class 8	8	2000	2000	2000	2000	1999.99	1997.57	1594.56	0

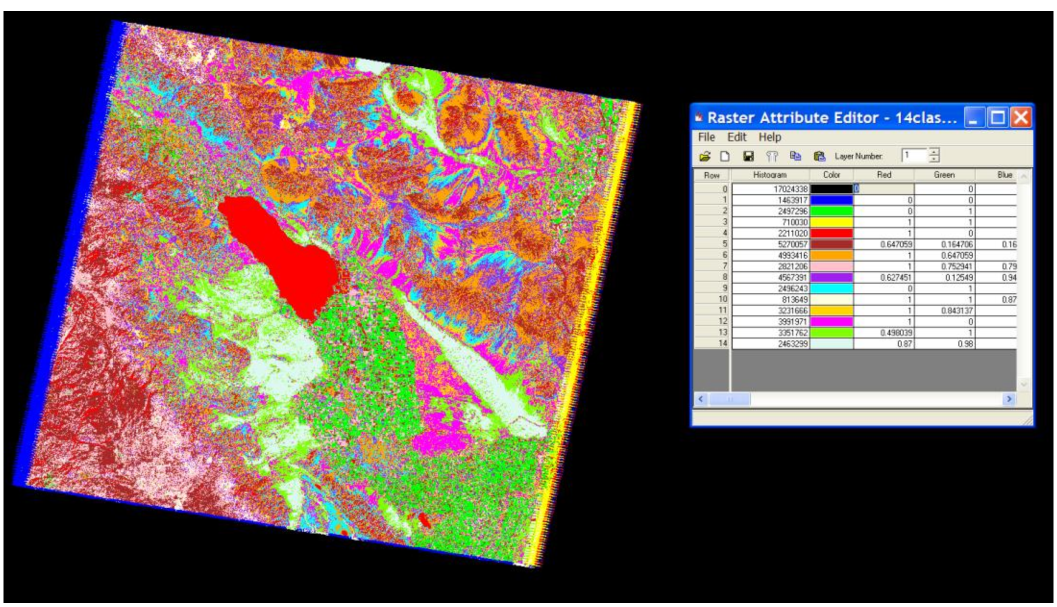
Close

From lecture, 2000 indicates perfect class seperability, and anything below 1800 indicates spectral confusion. Across the 14 classifications, water was the only separable class. There was a definite upward trend in spectral confusion as classes were added to unsupervised classifications. The spectral confusion also became finer-grained as classes increased. On the other hand, as classes increased, certain sets of spectral classes increased in seperability.

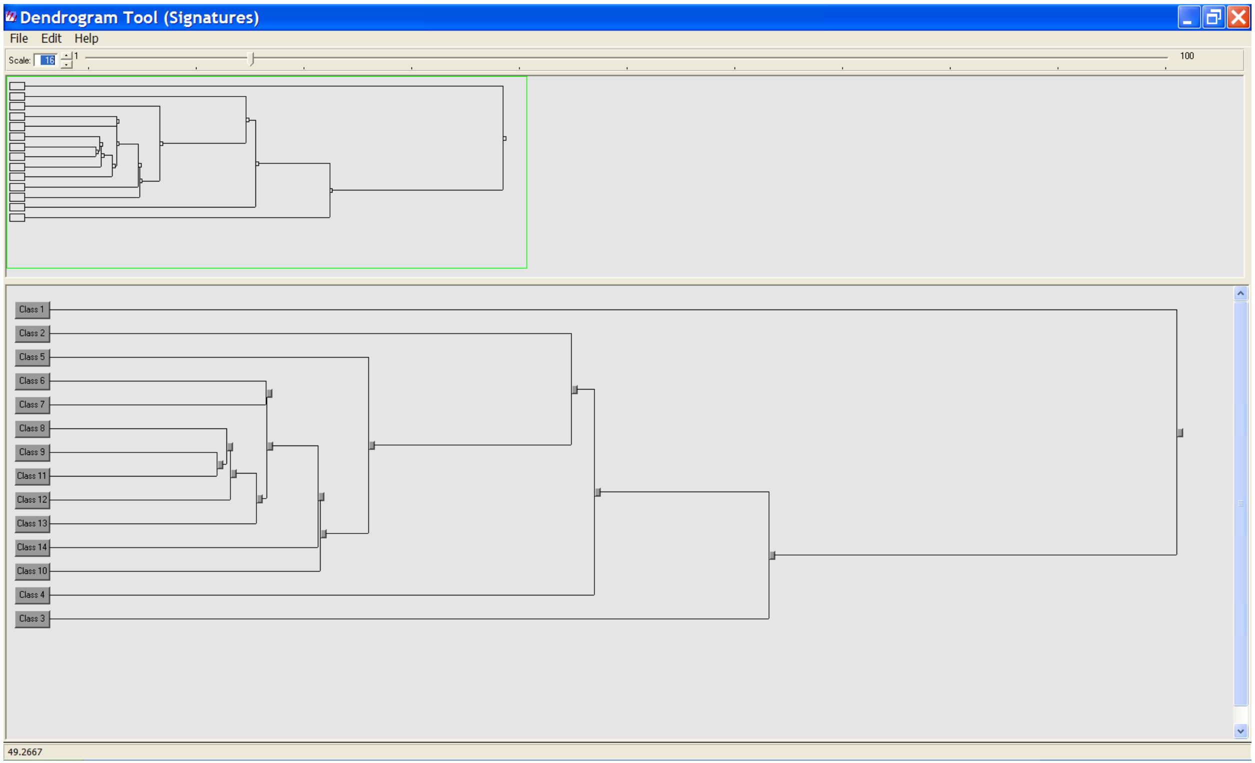
7 class unsupervised classification, colored to show class confusion



14 class unsupervised classification, colored to show class confusion



In the above images, these trends are noticeable. I had change the color scheme of the classes to visually highlight the spectral confusion I saw in the transformed divergence reports. In the 7 class image, mountain woodland and valley cropland are one class (green). Urban areas and desert areas are all mixed. In the 14 class image, cropland and mountain woodland have separated into separate classes, but urban and desert areas are still spectrally mixed. This is a trend that is constant in all 14 unsupervised classifications.



➔ 5:7(1688), 5:6(1761), 6:8(1533), 8:9(1432), 9:11(1379), 11:12(1521), 12:13(1602), 13:14(1542).

The above is a dendrogram from the 14 class unsupervised classification, and stats from the transformed divergence report (classes and measure). In the dendrogram, the most spectrally confused classes are in the middle, the more separable classes to the outside. There are different ways to represent a dendrogram – the one represented is Euclidean distance and simple agglomeration. It's not as easy to understand as a transformed divergence report.

By comparing increasing classes from unsupervised classifications, it became clear that more classes equate to higher accuracy. Certainly, comparing transformed divergence reports and dendrograms can help an analyst to define trends that may escape visual analysis. While looking over the 30 class image, I couldn't visually find spectral confusion except for urban areas. Looking over the (very large) transformed divergence report showed areas of spectral confusion that I couldn't see in the range of 30 colors.

The spectral confusion between urban and desert areas was unexpected. Oddly, some of the desert spectral classes seemed to be based on elevation. Perhaps the main building materials in urban areas are close to the composition of the surrounding mid-elevation desert ranges.

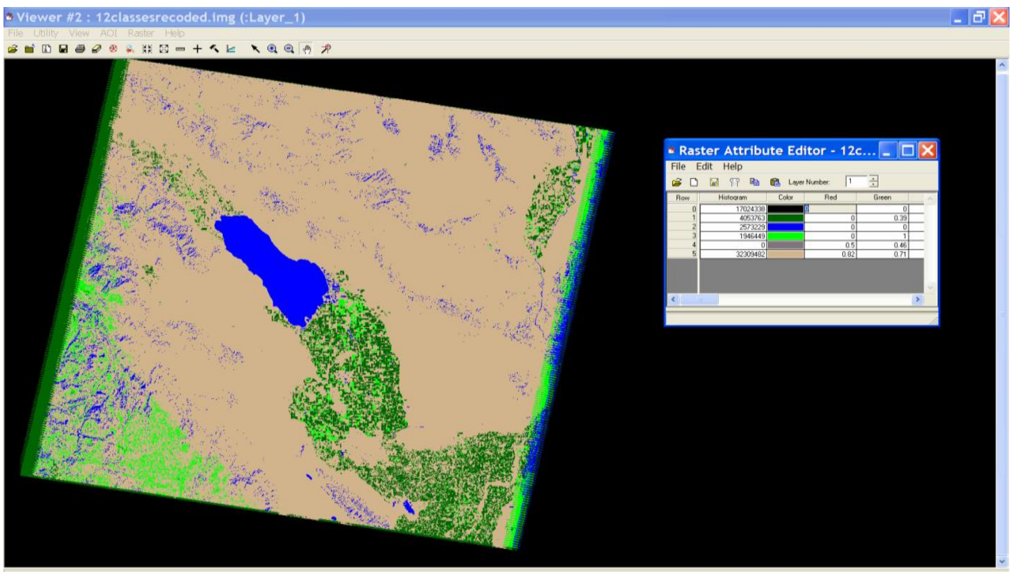
Part 2: Land cover classification

After I had run several unsupervised classifications, I was curious how a simple land cover scheme would be affected by increasing spectral signatures. Referring to the CalFire MLCD overlay, I came up with a 6 class land cover scheme; cropland, water, urban, woodland/grassland, desert, and barren. However, it became clear that barren land would be difficult to represent. I modified the scheme to 5 classes, subsuming barren land under the desert class. The land cover scheme alteration happened after I had run several unsupervised classifications as the basis for my outputs – I decided not to rerun them to match the altered scheme as I didn't have enough time.

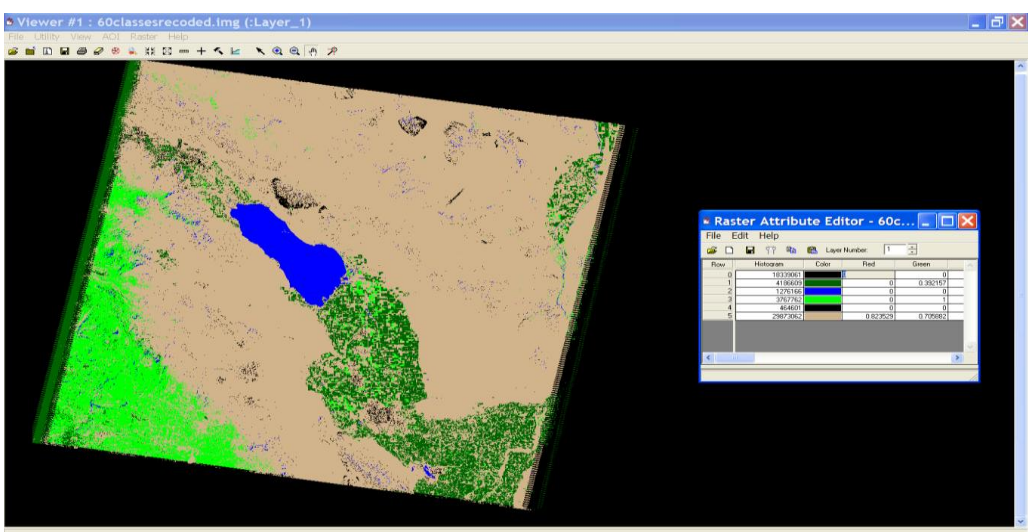
I used the recode tool on 10 separate unsupervised classifications, producing 10 land cover classifications. I had originally planned to start with 6 signatures, and increase by multiples of 6 until I reached 60 signatures (6, 12, 18…60). These were meant to match the original 6 class scheme, but instead were classed under an altered 5 class scheme. I'm not sure how this affected my results.

Using the recode tool was very easy. By turning signature opacity on and off on the unsupervised classification input, I was able to visually asses what information class to place each signature in. It got a little complicated once over 30 signatures. I was not able to assign an equal number of signatures to any of the information classes in any of the recodes.

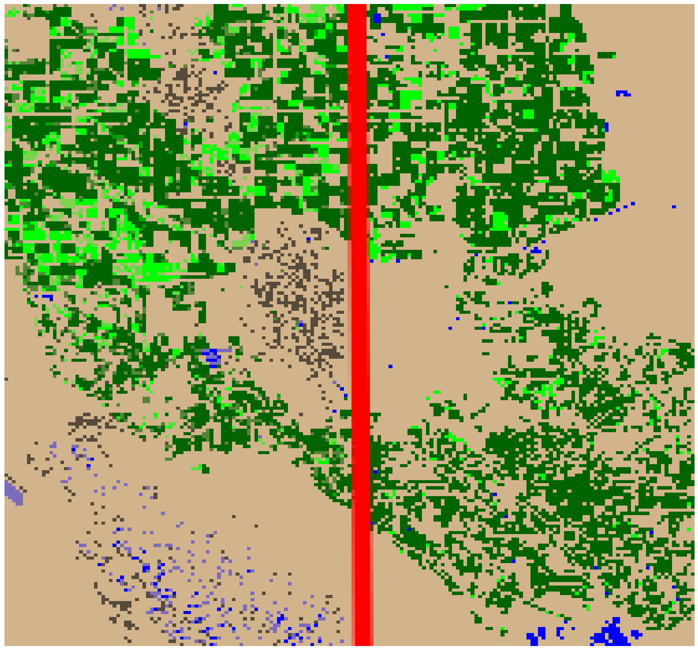
12 signature recode



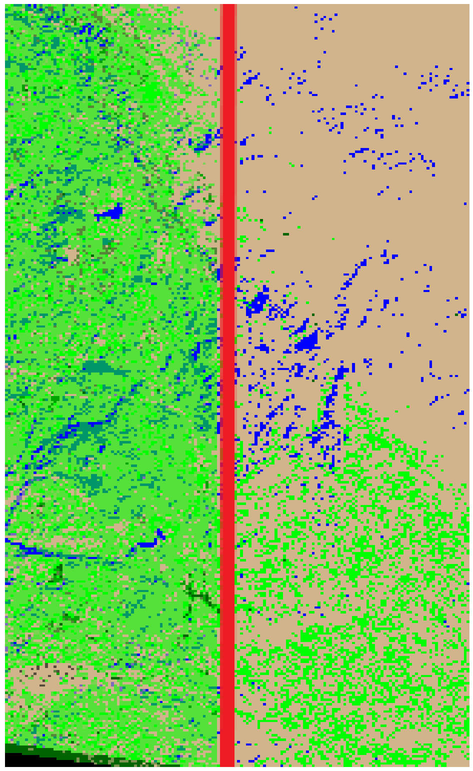
60 signature recode



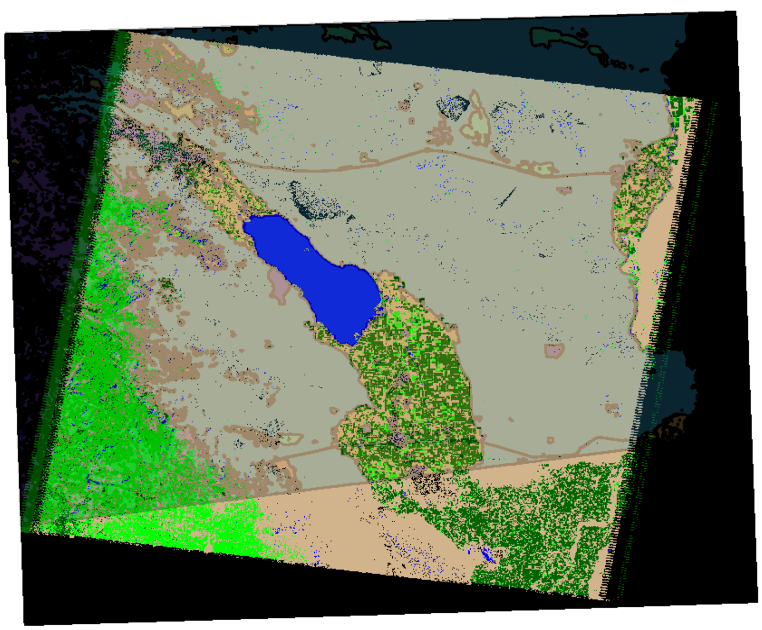
Each of the resulting 10 images was colored in the same way to help with visual analysis. During the recoding process, the desert information class was assigned the most signatures. The urban class was assigned the least amount of signatures. I used the swipe tool in Erdas to compare a few of the images after they were all colored in a standardized scheme, and a few interesting results emerged.



The above is a comparison of the 60 signature recode (left) with the 12 signature recode (right). The image area represents the location of El Centro, county seat of Imperial County. Urban areas were the hardest to assign signatures to in any of the recoded images. El Centro only barely shows up with 60 signatures to work with. The urban and desert spectral confusion seen in the first part of my project is again shown here. Even though it is spectrally absent from the 12 signature recode on the right, the boundaries of El Centro can still be inferred from boundaries of the cropland (if you knew the city was supposed to be there).



The above image is another comparison of the 60 signature recode (left) with the 12 signature recode (right). The image represents the start of the vegetation in the mountains between the Imperial Valley and San Diego. The vegetation level is more accurately represented by the 60 signature recode.



The above image is the CalFire MLCD as a 75% transparent overlay layered over the 60 signature recode. The CalFire overlay stops at the border, but it is still useful for comparison of the image on the Mexico side; the terrain is mostly similar. The recode of an unsupervised classification came close to matching the boundaries defined in the overlay. The more signatures that are assigned to an information class, the more accurate the image output will be. If I were to do this over, I would collect more signatures and some ancillary data on elevation, crop types, and urban boundaries. Using recode on an unsupervised classification is a good way to create a basic land cover classification, but the inaccuracies make it suitable for basic analysis only.