Lab 7

Change Detection in Urban Landcover Using Supervised Classification

Theodore Lessman
GEOG 485 F17 Kohler

Geographic Extent:

The focus of this lab is the area surrounding northern Deschutes County, in Eastern Oregon. The specific region includes the area between latitudes 43.909 and 44.468 and longitudes -120.634 and -122.002. This area includes the cities of Bend, Redmond, and Prineville and other smaller neighboring towns such as Alfalfa, Tumalo and Powell Butte. The western extent also includes the features in the Cascade Mountain Range: Mt. Bachelor, North and South Sister and Broken Top. The Deschutes River runs through the central portion of the extent while Prineville Reservoir and Crooked River reside in the eastern portion of the extent.

Data Source:

Landsat 5 Reflective Imagery (Thematic Mapper)

-Bend_OR/2011/LT50450292011204PAC01 (2011),

-Bend_OR/1985/LT50450291985228PAC09 (1985)

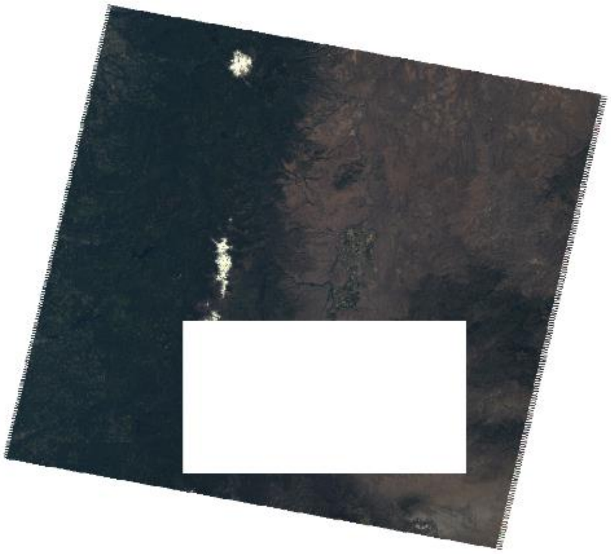
The imagery has a spatial resolution *or pixel size (one edge)* of 30 m, (band 6 resampled from 120m) a spectral resolution across 7 EMR bands including 0.45-0.52μm (1:blue), 0.52-0.60μm (2:green), 0.63-0.69μm (3:red), 0.76-0.90μm (4:NIR), 1.55-1.75μm (5:SWIR1), 2.08-2.35μm (7:SWIR2), and 10.40-12.50μm (thermal). The data has a radiometric resolution of 2^8, represented by digital numbers (DN) 0-255. The imagery was captured on August 16, 1985 and July 23, 2011.

Overview:

Human development of the Earth's surface has occurred throughout time; however, the last century has seen the greatest transformation of our world with industrialization. An increase in urbanization is has contributed to the spread of the built-up environment as more people live and work in cities. By implementing urban growth boundaries, Oregon has slowed the sprawl of urban development, though infilling and non-urban built infrastructure still can still increase the amount of developed land. By performing a supervised classification on both datasets included un this analysis, using training samples from the multispectral imagery, we can use machine learning techniques to quantify the amount of change in built-up land which occurred in the Deschutes county area between 1985 and 2011.

Pre-processing:

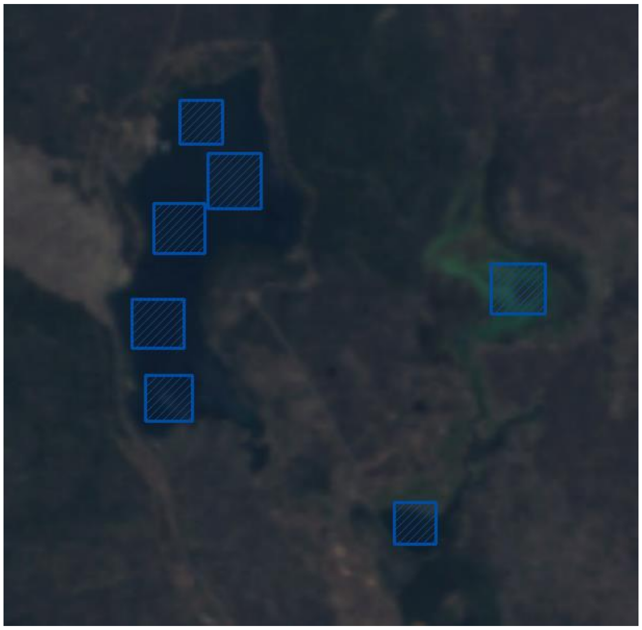
Before outlining any training samples or applying any classification algorithms, the images needed to be made into a raster catalogue (.vrt) so that it could be displayed as a composite image. Additionally, to convert from Digital Number (DN) to Surface Reflectance, I used the QGIS Semi-Automatic Classification Plugin to apply the formula using inputs derived from the metadata files (path radiance, viewing and illumination direction atmospheric transference, downwelling diffuse irradiance). Finally, I set the processing extent in ArcMap for my intended study area in order to reduce processing time. Adjustments to contrast and brightness were required for image display and interpretation.



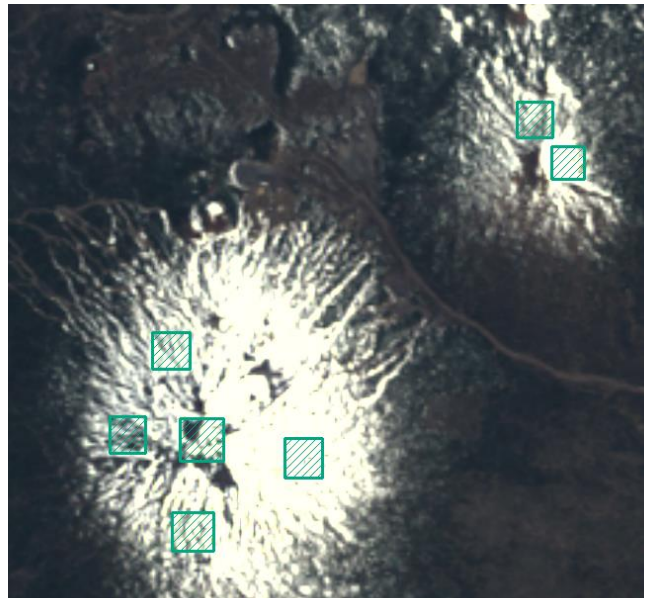
Processing extent of imagery masked in white

Analysis:

For the classification I performed a Maximum Likelihood Classification within ArcMap. Training Samples were chosen to be comprised of areas greater than 100 sq. pixels, where they were homogeneous and representative of the variation that occurs across the extent for each landcover type. I began with the 2011 imagery and selected training samples of water. Due to size constraints of training data, selection of sites was limited to reservoirs and lakes. After a successful classification of water/nonwatery, I progressed through each class, testing the classifier along the way. For the 1985 imagery, I retained and reviewed the 2011 training sample placements, and removed those which did not meet landcover type or uniformity. Additional placements were made within the 1985 set and refinements were made to attempt a better classification due to change in landcover (lack presence of snow, or built infrastructure, etc.).



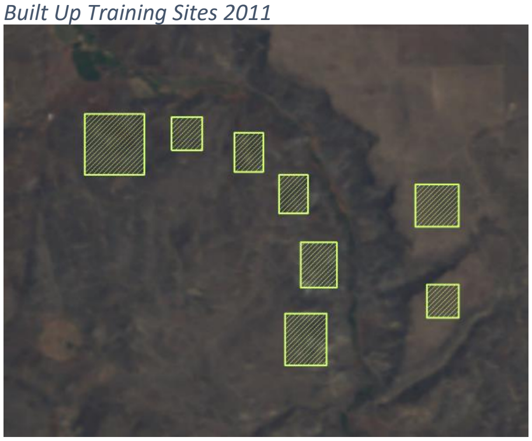
Water Training Sites 2011



Snow/Ice Training Sites 2011

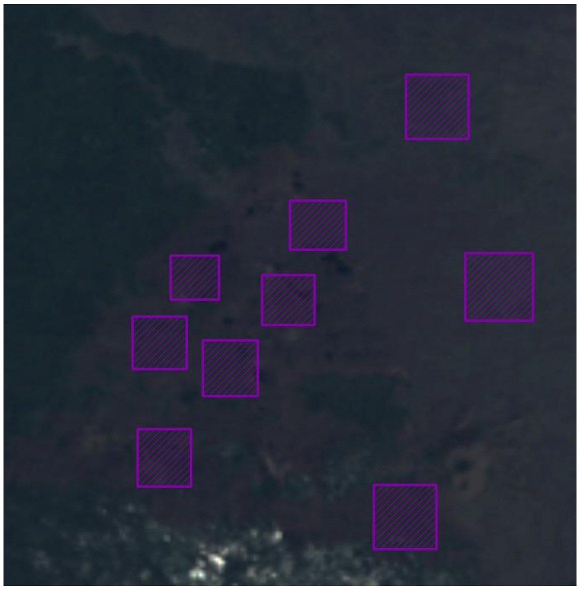
Built Up Training Sites 2011

Agriculture Training Sites 2011



Grassland Training Sites 2011

Forest Training Sites 2011



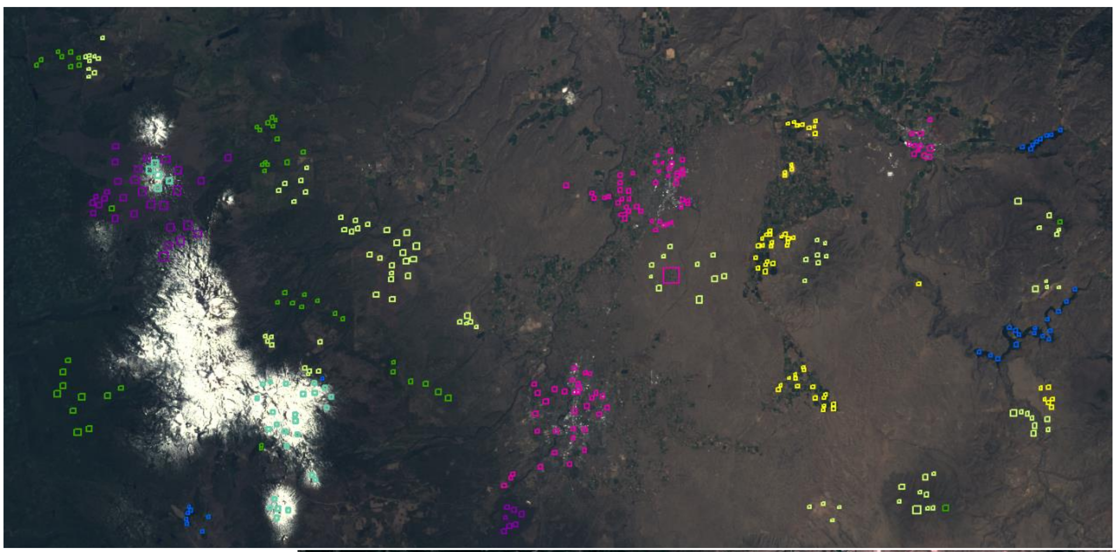
Barren Training Sites 2011

ID	Class Name	Value	Color	Count
1	Built Up	1		18226
2	Agriculture	2		6538
3	Grassland	3		18570
4	Forest	4		9831
5	Water	5		3840
6	Barren	7		15059
7	Snow / Ice	9		6611

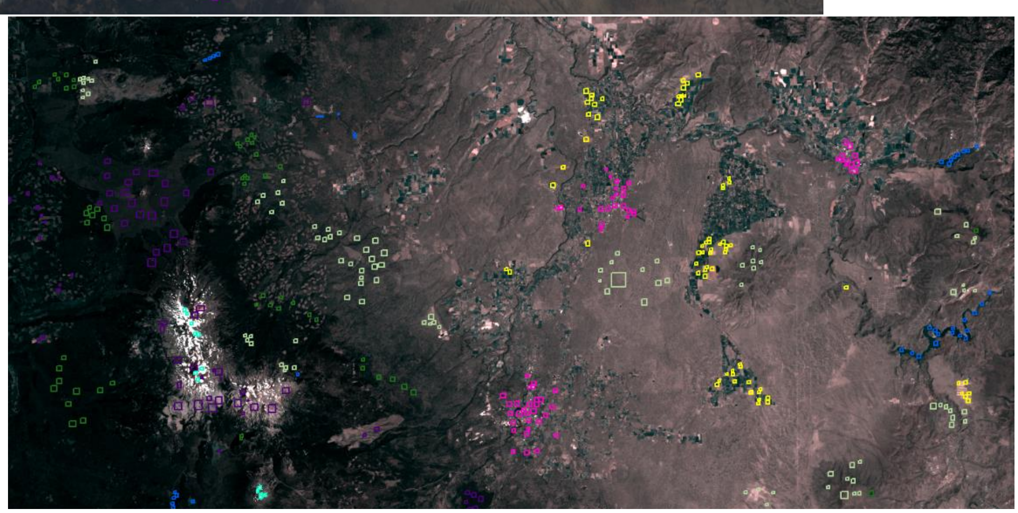
2011 Classes

ID	Class Name	Value	Color	Count
1	Built Up	1		12629
2	Agriculture	2		10068
3	Grassland	3		21269
4	Forest	4		11574
5	Water	5		4539
6	Barren	7		26169
7	Snow / Ice	9		1591

1985 Classes

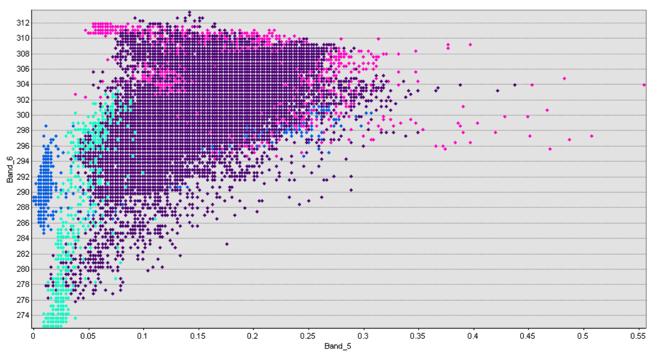


2011 All Sites

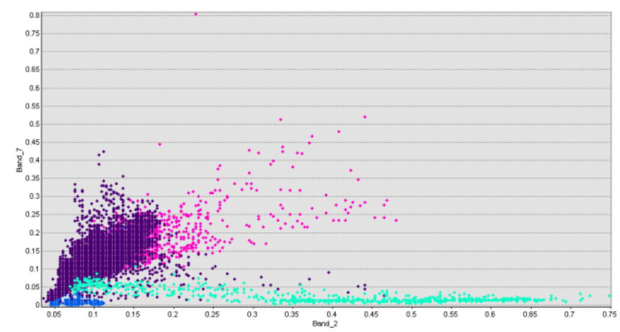


1985 All Sites

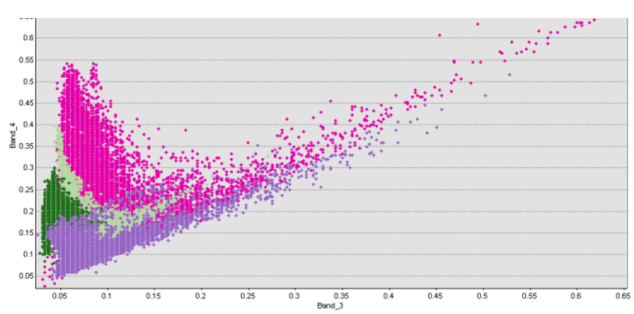
Spectrally, the 2011 dataset was better differentiated than the 1985 set. (histograms were unavailable). In the scatterplots for 1985, urban and barren were highly spectrally confused, and despite my efforts I was unable to get a better classification out of the two. Water and snow also shared this wavelength space to an extent, however the pair of bands 5x6 and 2x7 did a decent job at separating them. In the 2011 dataset, bands 4x5 did a remarkable job of separating each class. Here, rangeland, forest, and barren were the most spectrally confused, with some overlap. This is rectified to some degree looking at 4x6.



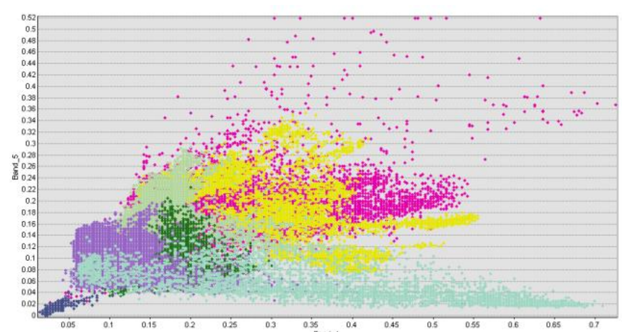
Bands 5,6 1985 Urban and Barren confused, Water and Snow separating.



Bands 2,7 1985 Urban and Barren confused, Water and Snow separating.



Bands 4,6 2011 Barren, Forest, Grassland separating.



Bands 4,5 2011 All classes are clustered, some overlap but discernable and predictable.

Classification of the 2011 dataset was excellent, with most pixels being assigned as expected. Interestingly some portions even included in training sites became classified as other landcovers, most notably urban golfcourses being classified as rangeland or agriculture (despite attempts to ensure built up categorization). The 1985 imagery classified well aside from in the mountains, where bare rock became classified as urban. As such, more urban exists than there was in reality, which will ultimately throw off my analysis considerably.

Using raster calculator, I took a difference of the 2011 classification and the 1985 classification to see where the change occurred. Value was assigned based upon the difference in Anderson Classification number (1,2,3,4,5,7,9). If there was no change, the result was given the value **0** (for instance, urban to urban: **1-1=0**), while the remaining values ranged from **[(-8),(-1)]** and **[1,8]** (for instance grassland to urban: **3-1=2**, or forest to snow: **4-9=(-5)**).

Finally, a change matrix was calculated using the “Tabulate Area” Tool, then converted from pixel count to actual square footage. To read the tables, find the column of landcover type for 1985, and find what it changed to for each row in 2011. For instance, no ice/snow became forest.

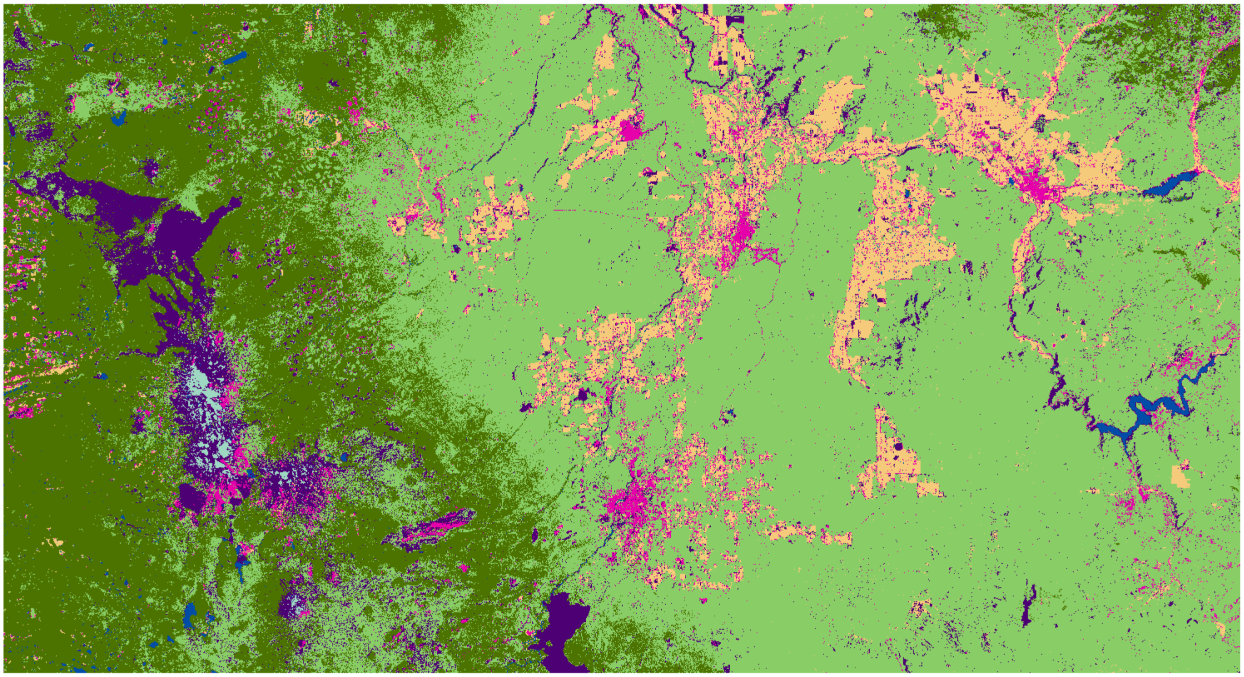
Looking at urban change, keeping in mind that a large amount of this is will have poor ground truth accuracy in 1985, urban land is listed as changed into the many different classes. This can be due to borders (such as rivers being classified as built land), spectral variation in samples (yards, golf courses), or erroneous classification (like the barren/urban spectral confusion in the mountains.

Looking at what became urban, we see that over 2.6 million sq. ft. or rangeland became urban (which makes sense given the growth which can be seen around existing cities. Likewise, nearly 1.5 million sq. ft. of agriculture was made into built land (this could be due to infrastructure expansion, inbuilding, geographic expansion, or error in classification). Much of the erroneous urban land from 1985 became barren land in the 2011 map, as expected. On a positive note, only 122,760 sq. ft. of forested land is classified as urban, though this could include some error in training data definition of what counts as forested land. 23,070 sq. ft. of water is listed as becoming urban, which is likely due to river border error, and 4440 snow becoming urban is obviously just noise in the classification.

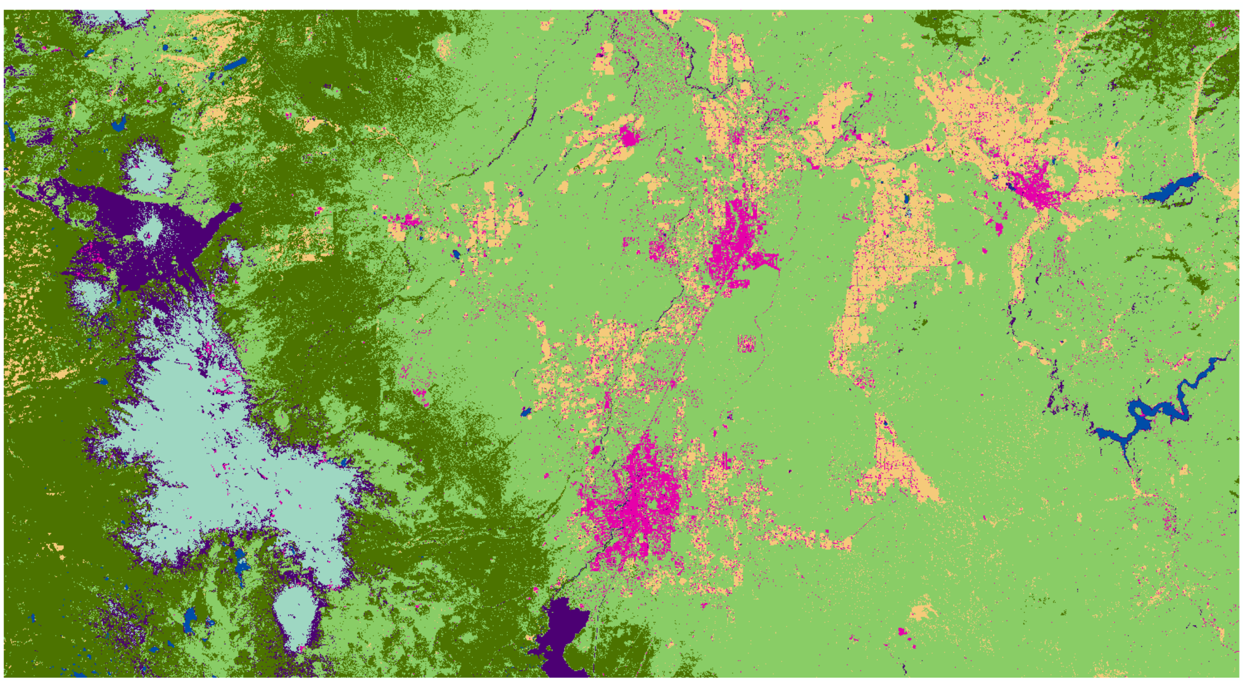
Conclusion:

Machine Learning Algorithms can vary greatly in complexity, and can be very powerful tools, however, they are only as accurate as the data they are fed. Data needs to be properly formatted, in expected domains of values for your specific tool, and training data needs to be representative of the dataset you are working with. This classification could be improved by obtaining data with less seasonal variation (despite the similar date range, yearly seasonal difference can vary widely), and most notably by better defining the 1985 dataset's training data. Had I built the training set from the ground up, perhaps the overlap in barren and urban would have been less pronounced, and less inaccurate classifying would have occurred. Also, reducing my extent to ignore the mountain range would have eliminated much of the error that was introduced from the snow/barren border. Finally, having a better definition of each ground cover type, or using a pan chromatic layer for pansharpening, may have allowed me to better discern features to obtain more homogenous and accurately categorized training samples.

Classification Raster:

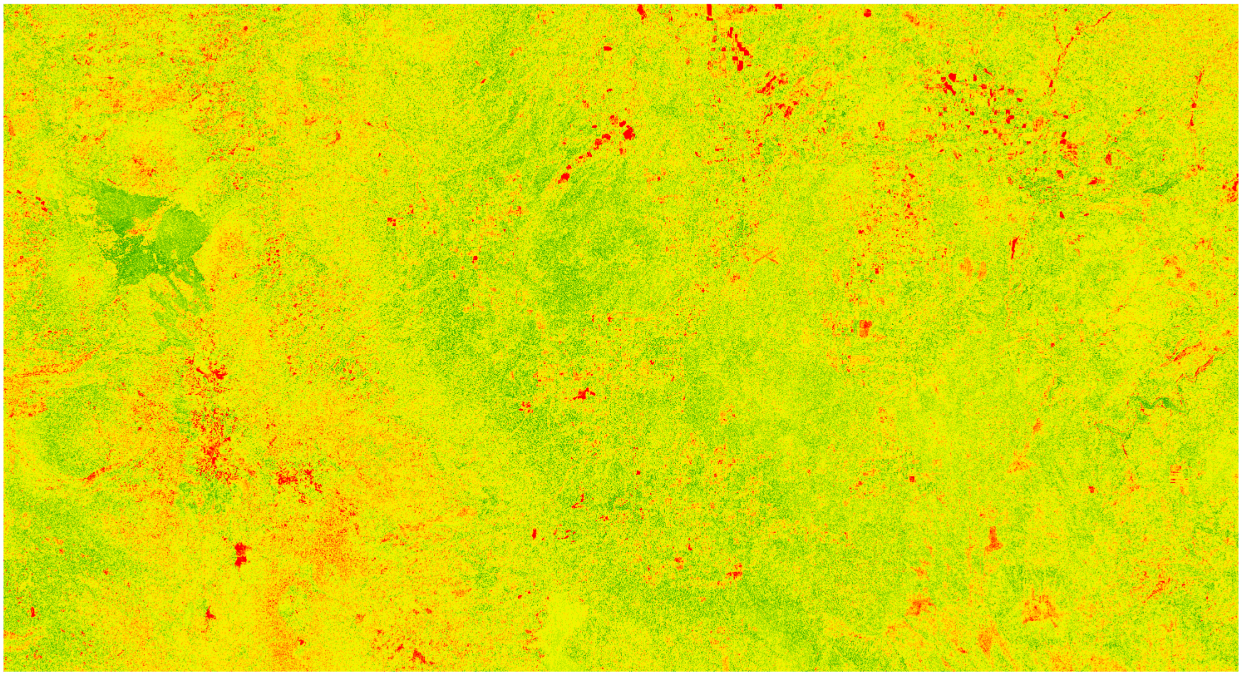


1985 Classification. Note erroneous urban presence in mountain barren land.

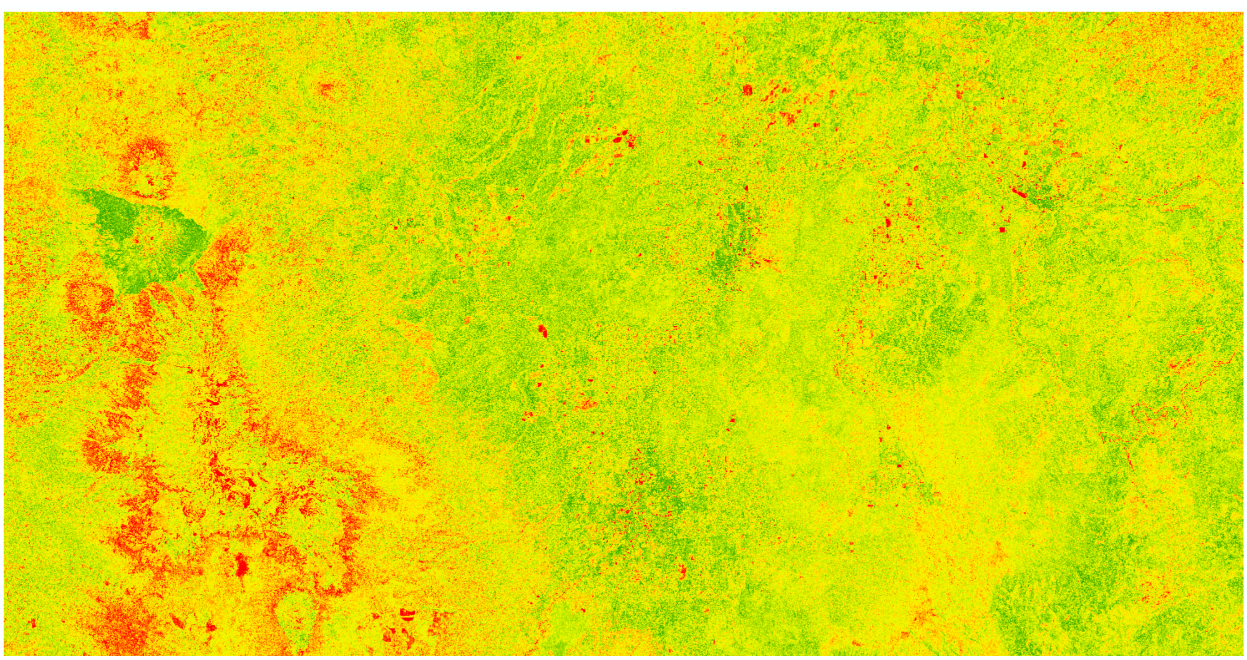


2011 Classification. Well defined urban areas and built land including highways.

Confidence Raster:

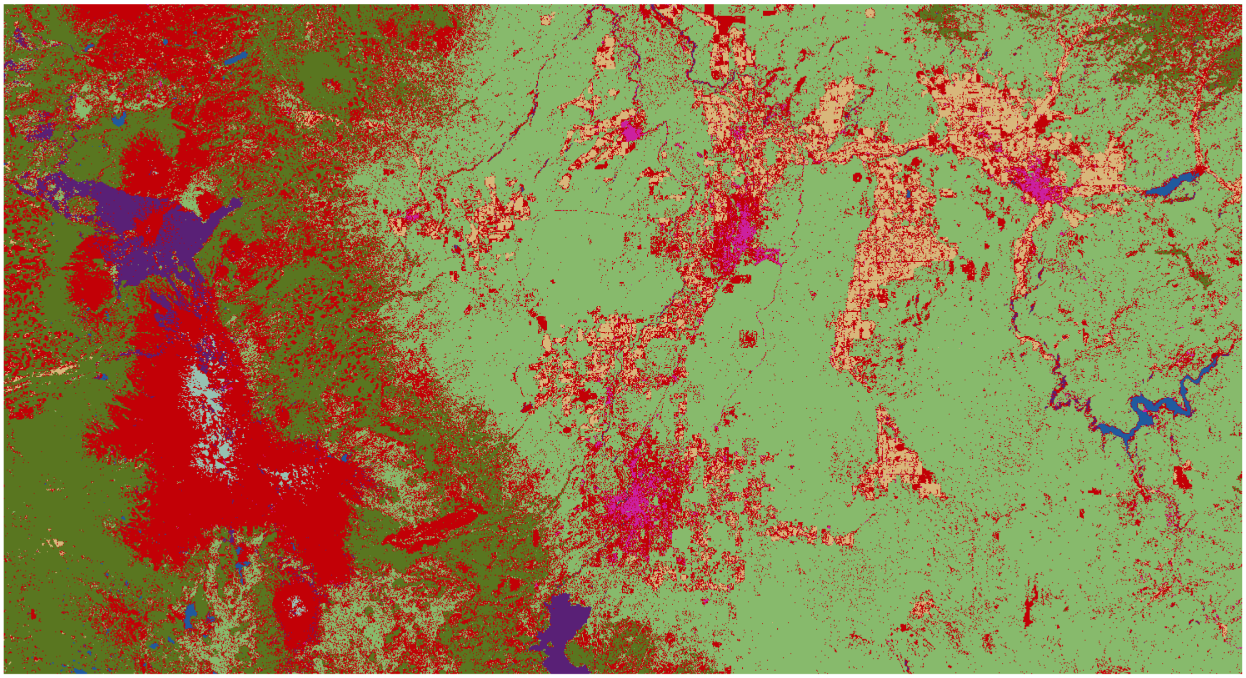


1985 Confidence Raster. Urban areas (including the incorrectly classified section), barren, and grassland generally are well defined in training samples.



2011 Confidence Raster. Very high confidence through most of extent. Lowest confidence in border regions, specifically around snow or water.

Change Detection Raster:



Change detection using raster calculator and map algebra of difference between 2011 and 1985. Red indicates that change occurred while other colors indicate specific classes where landcover remained the same. Most change occurred around cities, snow, and forest/grassland border.

Change Matrices:

Tabulate Area, raw pixel count

2011\1985	Urban	Agriculture	Rangeland	Forest	Water	Barren	Ice/Snow
Urban	56466900	44235900	78545700	3682800	692100	17279100	133200
Agriculture	41259600	311050800	96838200	34857900	2694600	25735500	9900
Rangeland	82809000	92332800	3290806800	287874000	5778900	120401100	1800
Forest	21257100	9820800	224538300	1119888900	8671500	33078600	0
Water	405900	526500	492300	342000	23018400	3850200	37800
Barren	3728700	1298700	29599200	94754700	3970800	129418200	185400
Ice/Snow	19434600	2254500	52868700	114465600	4326300	99042300	20669400

Values Corrected (in sq ft) for Grid Size (30m)

2011\1985	Urban	Agriculture	Rangeland	Forest	Water	Barren	Ice/Snow
Urban	1882230	1474530	2618190	122760	23070	575970	4440
Agriculture	1375320	10368360	3227940	1161930	89820	857850	330
Rangeland	2760300	3077760	1.1E+08	9595800	192630	4013370	60
Forest	708570	327360	7484610	37329630	289050	1102620	0
Water	13530	17550	16410	11400	767280	128340	1260
Barren	124290	43290	986640	3158490	132360	4313940	6180
Ice/Snow	647820	75150	1762290	3815520	144210	3301410	688980