Wieland

Geog 342

GEOG 342: Remote Sensing Final Project Documentation

Exploring Unsupervised Classification and Interactive Supervised Classification in Order to Characterize Impervious Cover

A. Summary

Isocluster Unsupervised Classification and Interactive supervised classification were used to classify levels of impervious cover in a certain spatial extent within a watershed. Percentage of impervious cover was calculated using raster geoprocessing techniques, and numbers were compared to an alternative method that does not use image classification techniques. Methods, technical issues, and results were documented.

B. Purpose: what I hope to accomplish

I hope to determine the levels of impervious cover (IC) in different spatial extents in a watershed by using multiple image classification techniques. I will run the particular type of image classification, and then measure within the specified polygons the amount of data that appeared as impervious cover. I will compare the results from these image classification techniques to a previous method we have used in our work group, in order to compare the two and hopefully, evaluate the overall accuracy of each of the image classifications.

C. Description of image processing tasks and methods used to create the outputs

Introduction

I work for the Office of Environmental Health Hazard Assessment (OEHHA), part of Cal/EPA, as a student intern. Most individuals working at OEHHA are toxicologists, and they evaluate human health in regards to many different risks. The group I work in is slightly different: I am in the ecotoxicology branch, and we do ecological risk assessments (ERAs). Our primary ERA focuses on watershed risk assessment and it is our goal to determine the best ways to run risk assessments for a species of concern (generally Salmon and Steelhead in our watershed). We characterize different types of risk by evaluating chemical and physical stressors, as well as their potential sources.

An important element of an ERA is by performing various types of landscape analysis to understand important characteristics of the watershed. It is common for scientists to use GIS to understand spatial relationships and land uses, among other things, in a watershed. One of the metrics our group measured in a watershed we have been studying is the amount of impervious cover. Levels of impervious cover have been proven to be significantly correlated to aquatic ecosystem health, so we wanted to understand some characteristics of our watershed in terms of urbanization and land disturbance.

Using GIS, we measured impervious cover at a variety of spatial scales, and ran correlations against biological data we had to try to understand the relationship between the two. The method we used to measure impervious cover was not by using remote sensing techniques. Instead, we used land use data that characterized parcels as having a land use designation (LUD). For example, most urban area LUDs are either residential, commercial, industry, or open space. Some rural ones are agricultural and rural residential. Over an intensive development and review period, OEHHA developed impervious surface coefficients (ISC) for California: the average amount of impervious cover associated with every different LUD. For example, it was determined that residential parcels are about 44% impervious cover: the driveway, house, garage, and shed in the backyard might take up that space, while the front and back yards might take up the other 56%. Open space is characterized as being 2% IC; heavy industrial is characterized as being 91% IC.

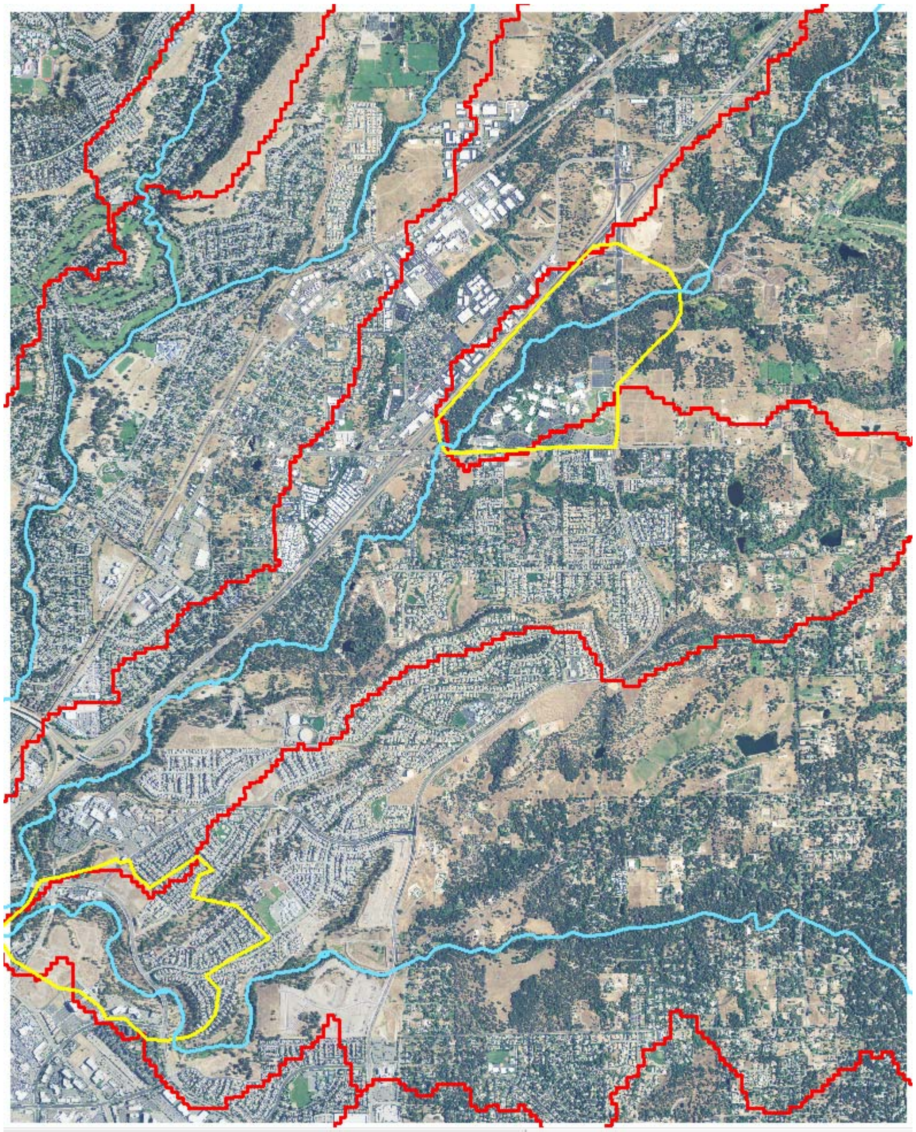
The method we have used is as follows:

- Correct for non-conforming land uses by changing the land use type to it's appropriate current use (example: an area could be zoned as commercial, but it has not been built on yet. If there is bare soil, we might call the area 'open space'.
- Measure the total area of each type (in acres) of land use within a given extent
- Apply the ISCs to the areas of all types of land use
- Sum up 'acres impervious' and divide by 'total acres' to determine percent impervious within any given extent.

We have a good amount of confidence in this method, but it takes a lot of work and analysis is slow. I am going to evaluate some image classification techniques in order to determine their ease of completion and their output results in comparison to the results obtained using the methods described above.

Acquiring Data

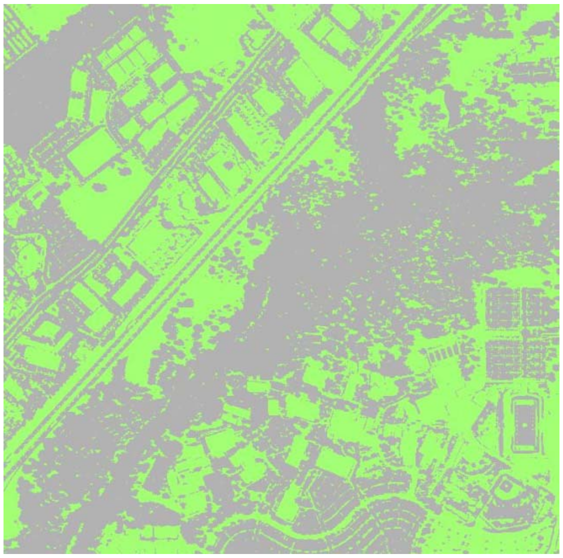
I wanted to download some aerial imagery of the watershed we were studying, so I went to CalAtlas.com and downloaded some Digital Orthophoto Quarter Quads (DOQQs) that are produced by USGS. Specifically, I downloaded the 2009 combined near-infared (NIR) National Agriculture Inventory Program (NAIP) DOQQs; because I was advised that this type of imagery had 4 bands and I could do more in terms of image processing. Pasted below is the DOQQ I did image processing on and the outlines of the streams and subwatershed boundaries that I plan to use in this analysis:



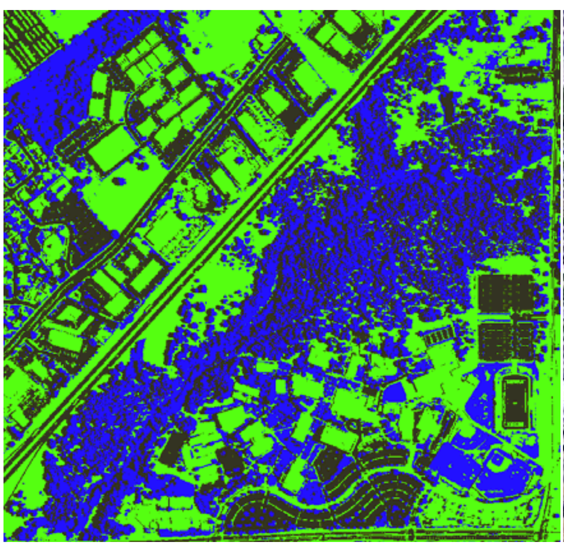
The blue lines are the streams, the red are the sub-watershed boundaries, and the yellow are the two polygons I intend on using to compare levels of impervious cover reported.

Analysis

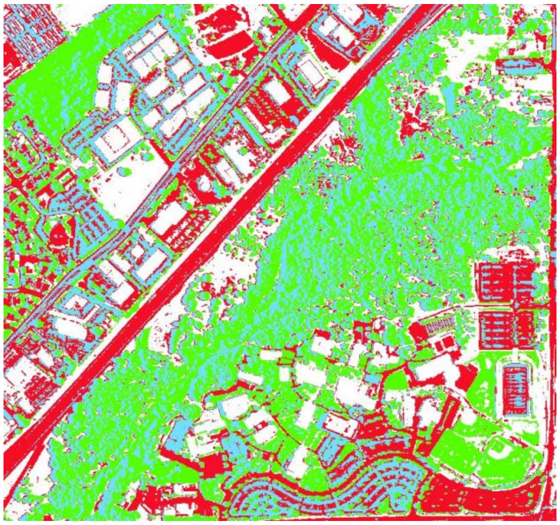
The first method I tried was the iso cluster unsupervised classification, which is a classification method that performs image classification on the input raster bands. First, I calculated statistics for my raster input dataset, which then allows me to perform image classification. Then, I ran the tool using 2 classes. The classification below resulted (left), with the original image (right).

These results show that some areas, like roads and parking lots were identified, but the tool does a poor job of characterizing impervious cover: putting most bare soil in the same class as hardscape, such as rooftops and a freeway. This was not going to be an accurate enough classification method, so I tried the same method using 3 classes. The results and comparison with the original image are below:

The results from using three classes look more promising: most green space like trees and the baseball field look to be accurately classified, and now much of the bare soil appears to be identified as well. However, the building tops that show up as white are still being put into the same class as the bare soil. To examine this analysis further, I did the same process with 4 classes. Here are the image results:

Here, the picture actually gets more distorted then using only 3 classes. The bare soil is still in the same class as the rooftops, and now some of the roads have been split up into two classes. I didn't think that any variation in classes for unsupervised classification would be adequate to measure against our impervious surface coefficient method, so I tried a different method of image classification.

The next technique I tried was Interactive Supervised Classification. This method requires me to draw multiple training samples for the pixels I want to characterize. Then, it will classify the image based on the training samples I have created. This technique differs from traditional supervised image classification in that I do not need to create a signature file (.gsg) in order to classify my image. Also, the tool works quickly as long as I have my pyramids built. In order to differentiate impervious cover from pervious cover, I created training polygons of areas that I thought characterized pervious cover such as trees and dirt, and impervious cover such as rooftops and roads:

In this application, I have created a 'pervious' polygon with about 30,000 pixels and a 'impervious' polygon with about 4,600 pixels. I have good confidence that both polygons accurately represent my two classes I'm interested in. After I run the interactive supervised classification, here is my result:

This is roughly the same area as before. It appears much of the impervious and pervious cover was classified accurately, but now I want to measure these ratios within a polygon I have used for the ISC calculations.

The next step for me was to overlay a buffer polygon on my image results and look at percent 'red' (or impervious cover) that resulted. But first, I had to convert raster features to polygon features so that I could take a measure of area. After much wading through technical issues (see section D) I found a way to convert my interactive supervised classification raster into polygon features. Once I made the conversion, I created a field called "area_sqmeters" to get the area of the polygons pervious and impervious. My results were as follows:

Classification type	**Area_sqmeters**
Pervious	1006094
Impervious	374502

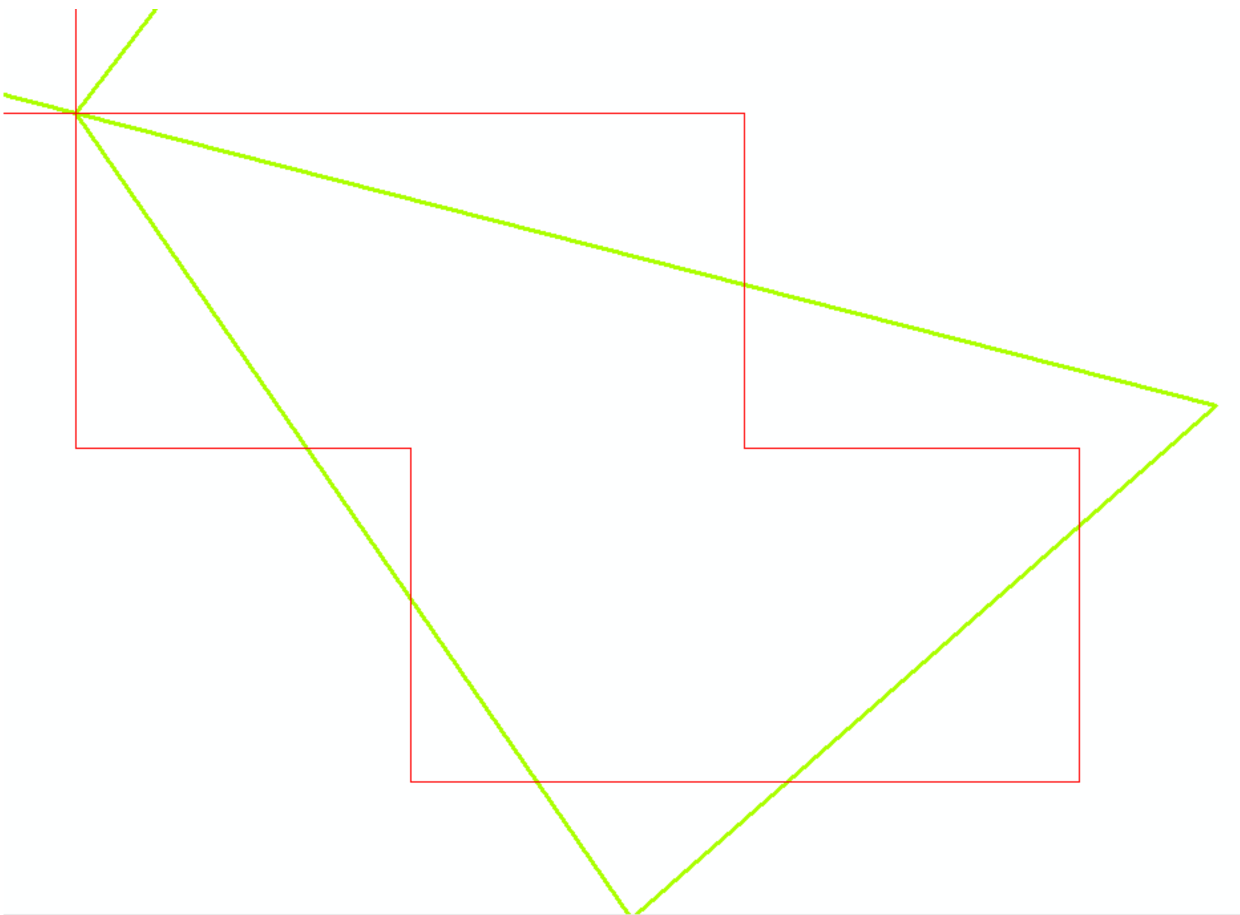
I know that I have to multiply this area field by meters squared in order to get total number of pixels that are classified as either pervious or impervious. So, after doing that, my numbers looked like this:

Classification type	Number of pixels
Pervious	10823961
Impervious	4029042

These results suggest that there is quite a bit more pervious land than impervious land in this polygon, which is what I expected to see. However, before I compared these numbers to my previous results on percent impervious cover, I wanted to verify their accuracy. There is a tool in Spatial Analyst called "Zonal Geometry as Table" that calculates multiple measures of geometry for each 'zone' in a raster, giving me numbers on number of pixels in my raster. The two zones I worked from are the same as the two classes. The results are as follows:

Zone (class)	Area
Pervious	10830614
Impervious	4030196

The numbers are very similar, but not exact. I spent some time investigating the raster data and the processing to try to figure out the reason for this. What I found was that I had missed an option in the 'raster to polygons' conversion box, which was to simplify the polygons. When this box is checked (the default), the raster polygon shapes will be smoothed to make them simpler. When it is not checked, every cell in the raster is calculated.



In the above picture, the red lines show the outline of the cell edges, while the green line shows the polygon that resulted from the simplify option. The areas do end up being similar, but larger changes are seen when thousands of these features are produced. With the 'simplify polygons' option unchecked, and converting the count from cell number to meters, my results were as follows:

Feature	**Cell number**	**Square feet**	**Zonal Output**
Pervious	1006201	10825112	10830614
Impervious	374419	4028149	4030196

These numbers are much closer to the output from the Zonal geoprocessing tool, but the results still were somewhat odd, since I measured the approximate area of the polygon I was working in with the measuring tool and I got a number closer to 1.3 million square meters, instead of 14 million. I then discovered that my mapping units were in square feet, which explained why the Zonal geoprocessing tool used square feet. After doing the math, I found this was the case.

Now, I was ready to calculate impervious cover within this polygon and compare my number to the alternative number we had. My interactive supervised classification yielded a result of 27% impervious cover within this polygon, and our method of using impervious surface coefficients had the same buffer as being 32% impervious. The numbers are pretty close to one another, which was certainly a promising result. However, a 5% difference in impervious cover could have large implications, so it would be necessary to investigate whether the interactive supervised classification had more error or whether the impervious surface coefficients had more error.

D. Discussion on difficulties/issues

One of the primary problems I had during this analysis was doing any type of geoprocessing on raster data. I do not have a lot of experience using ArcToolbox tools on input raster datasets, so when I received an error message (and I often did during processing) I spent a large amount of time trying to figure out first exactly what the error meant, and second what the solution to the error was. One of the first problems I came across was during my attempt to convert my raster features (from my interactive supervised classification) to polygon features for measuring percentages; I got an error that didn't make any sense to me. After researching online extensively, I found that I could not have any spaces in any part of my file path name and that was causing an error. Another error I made was a common one made by many GIS users and I was somewhat embarrassed to have made it: my vector data was in a different projected coordinate system than my downloaded raster data. I came to this realization when I tried to clip my raster data using a vector polygon, and the analysis did not produce results. I projected my raster in the appropriate coordinate system and the clip worked. Finally, my zonal summary reported my results in feet (my mapping units), and my input raster was measured in meters, which gave me odd summary numbers to compare to, but everything made sense once I did the conversion.

E. Conclusions/inferences: was outcome expected? Why / why not?

Regarding isocluster unsupervised classification, the method seems to have some strengths, it just depends on the type of image you're working on and what you want the end result to tell you. I was actually pretty surprised that the unsupervised classification gave me results as good as they were, and if the tool had characterized bare soil along with trees and grass, I might have used it to convert to vector data and calculate percent impervious. However, the interactive supervised classification was much more accurate and was a pretty fast process one I learned the method. So, I would say that as a first run, I was not expecting the outcome for classification of impervious surface to be so accurate. Basically, classification occurred just based off of the two training samples that I created. What I thought I might need to do was create multiple training samples, perhaps 3 for pervious and 3 for impervious, in locations all over my raster, in order to accurately characterize the varying pixel shades all over the map. However, this was not the case. My pervious sample basically had some light and dark trees and some soil, while my impervious sample had some building tops and roads. However, I'm thinking the interactive supervised classification assigned all pixels in the entire image to the classes that they most closely conformed to. That is to say, within the entire range of all pixel values possible, the range was split up into two sections (pervious training sample and impervious training sample) and all pixels were assigned the appropriate value. If I wanted to discern between roads and building tops, I would have to make more training samples and I hypothesize that it would be more difficult to discern between features that are in a much more narrow range of pixel values. For example, within two roads having different shades of color, one might get classified as a building because it is darker and more similar to the color shade of the building top from which I created my sample. I would likely want to correct for errors and run my supervised classification multiple times in order to get the most accurate results.

F. What did I learn?

I've learned about some of the functionality of image classification and the potential it can have for performing analysis on raster datasets. I've learned about how challenging raster analysis can be without a comprehensive understanding of rasters, and some of the raster geoprocessing tools, especially in Spatial Analyst. I've learned that interactive supervised classification is a good image classification technique that can be implemented relatively quickly and can produce results that are quite accurate, although I would like to evaluate this tool in more depth before I make accuracy comparisons between it and the impervious surface coefficient method.