Andrea Eibergen
Remote Sensing 371

Lab 8b Report

Introduction:

The goal of this lab is to create a land cover map of the greater Madison, Wisconsin area using medium resolution satellite imagery. Specifically, the goal is to create a raster representation of land cover classes from a band sequential image file. To do this, one must aggregate data into clusters, which we will do via supervised classification using support vector machines (SVM). Supervised classification, as defined in our lecture, is a classification technique wherein an algorithm separates the spectral signatures of the image into groups based upon the training data and classes provided by the analyst.

Methods:

The first step in supervised classification is choosing the desired land cover classes. For the image of Madison, the classes I chose were forest, urban, bare ground, water, and vegetated farmland. I chose this classification based off prior knowledge of the greater Madison area, as well as imagery from Google Earth. Next, I began selecting training sites that give the algorithm examples of pixels for each of the desired classes by creating "regions of interest". My strategy for selecting training sites was choosing very small areas throughout the map. I had originally tried using larger areas, but found that this created some confusion between classes – for example, if one selects the whole isthmus as their training site, it will include both urban areas and greenspace that could be confused with farmland. As a result, I chose individual pixels in areas that, based off the supporting imagery and previous knowledge, I thought fit my classification scheme.

For my first iteration, I classified my map using a maximum likelihood algorithm. The maximum likelihood algorithm considers the covariance, or relational variance, for each pixel that is being considered in the image. However, this algorithm assumes that the distribution of the samples is normal, so if the data is not normally distributed errors are introduced from the beginning. It takes most variables into consideration, but the disadvantages are that it is a non-spatial method of classification and it is computationally intense, which can be problematic depending upon the computer in use. One issue with this method is that if you do not provide a large enough breadth of pixels for each class, or if you do not have enough classes, you will end up with a large number of unclassified pixels in your image, which will not help you with your analysis.

Then, in my second iteration, I used the support vector machine classification algorithm to classify my pixels. The support vector machine generates a possible simple boundary between classes by transforming the feature space – this boundary separates the classes with the maximum possible margin for a line through the data. The input space is then transferred into the feature space (i.e. the line is straightened and the points are modified), which is then transferred back into the input space to create classes. One issue with this is that the output image is usually very speckled with pixels of differing classes in areas that did not make sense in the context of the map – for example, an urban pixel showing up in the middle of an area classified as farmland. This happened throughout my map, so I applied filters to attempt to rectify this.

The speckled pixels problem was solved through post-processing of my map – applying a majority filter or a sieve to the SVM classified map to rectify any over classification of the map. The majority filter is a three by three square that parses through the map. As it moves through the classified map, the filter compares each pixel to its neighbors and chooses how to classify the center pixel in the three by three square based off whichever class label is the most prevalent. Although this can be a good method, it resulted in making my map look quite clumpy and generalized, so I applied a sieve filter to my original classified map to compare results. A sieve filter is similar to a majority filter, but it allows the analyst more control. In a sieve filter, a three by three square moves through the map, but this time instead of choosing pixel classes based off the majority, it uses a number supplied by the user. For example, in my sieve filter, I tried using values of two, four, and five. I was most pleased with the result from using four, as it was a good median between the sieve filter finding too many or too few "problem" pixels. The final step with the sieve filter, as it does not apply class labels to the identified pixels, was to create mask using the problem pixels and then relabel them based off the majority filter map.

Results:

One difference that is evident is the large amount of land in the southeast portion of the image classified as bare ground in the maximum likelihood classification (Figure 1) that is classified as a mix of bare ground and vegetated farmland in the support vector machine classification (Figure 3). In the original image and other supporting imagery such as Google Earth, these appear to be areas that are a mix of bare soil and sparse vegetation, suggesting that these areas were either recently planted or recently harvested. The reason that the SVM image has more vegetated farmland in this southeastern corner is likely due to the filters used on the image. Since the data was compared to other pixels using a nearest neighbor approach, these areas were likely mostly vegetated and then the bare ground pixels were classified as vegetated due to their neighbors. Another point of interest can be found in Figures 2 and 4, which are the class area statistics tables. One can see that the maximum likelihood classification includes unclassified pixels, which are labeled as having DN numbers equal to zero, whereas the support vector machine has no unclassified pixels. This can be due to not providing enough breadth of pixels for each class or enough classes.

When comparing the class area statistics of these two images, it is evident that while urban (DN: 1), forest (DN: 4), and water (DN: 5) did not change very much between the two classification algorithms. However, vegetated farmland (DN: 2) and bare ground (DN: 3) varied between the two, with many values that were bare ground in the maximum likelihood classification being reclassified as vegetated farmland in the support vector machine classification.

Andrea Eibergen
Remote Sensing 371

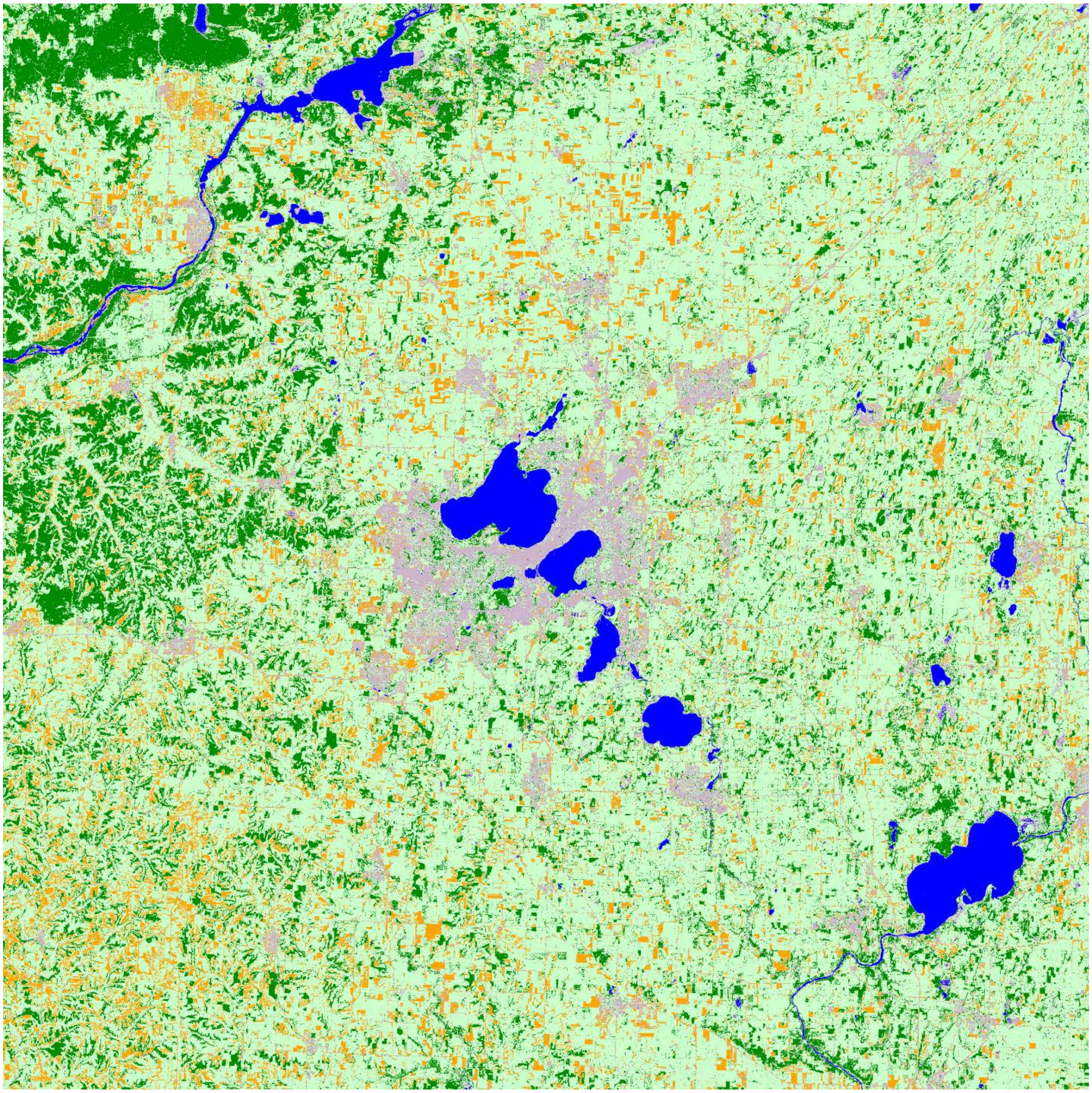

Figure 1: This map of the Greater Madison WI area was produced using the maximum likelihood algorithm, a supervised classification procedure. In the map, water is shown in blue, urban areas are light purple, vegetated farmland is light green, bare ground is orange, and forested areas are dark green.

DN	Count	Total	Percent
0	903	903	0.010695
1	653693	654596	7.742086
2	5522236	6176832	65.403222
3	759591	6936423	8.996301
4	1285412	8221835	15.223921
5	221535	8443370	2.623775

Figure 2: Class Area Statistics for the Madison image classified using the maximum likelihood algorithm.

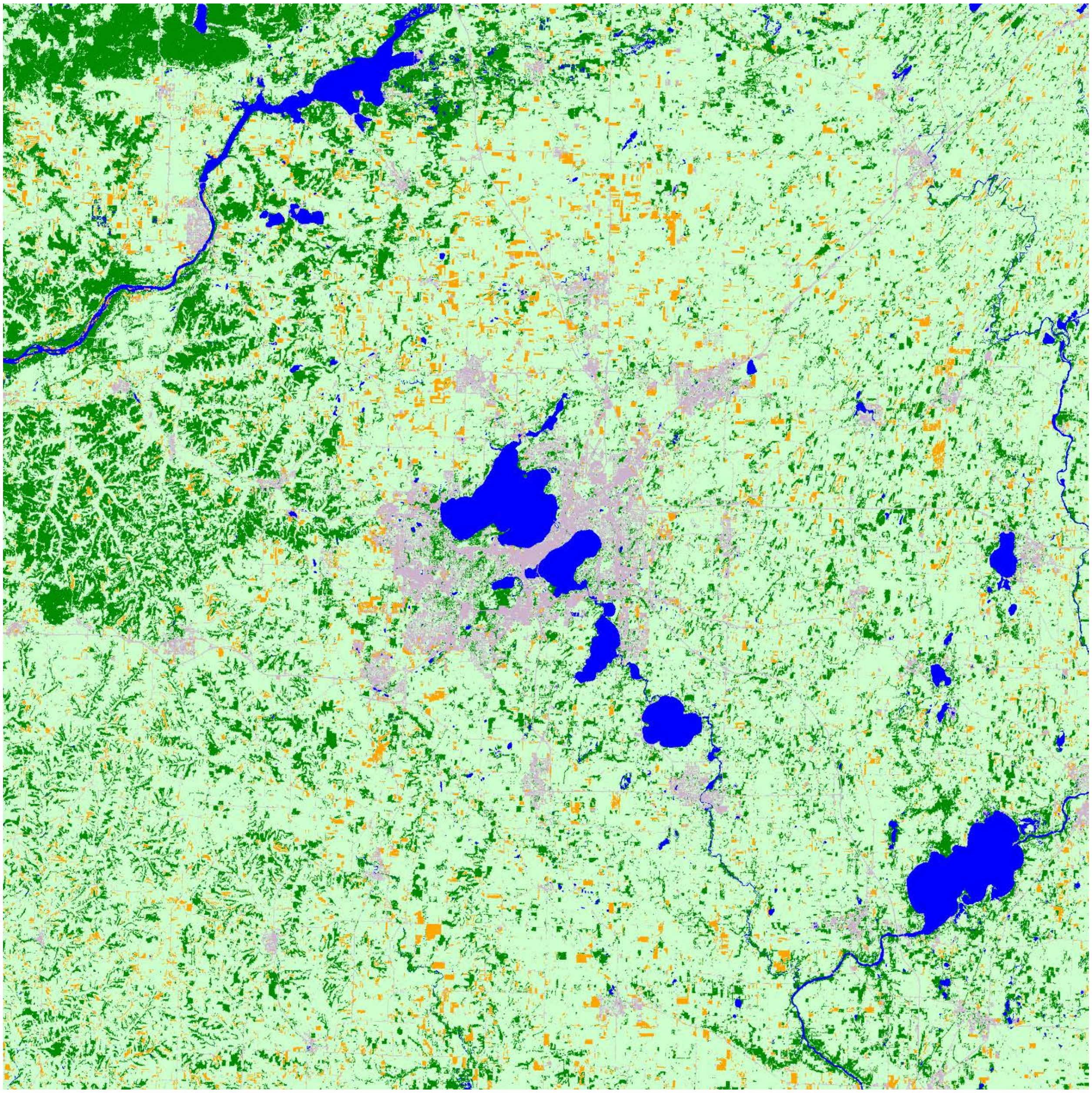

Figure 3: This map of the Greater Madison WI area was produced using the support vector machine classification algorithm, a supervised classification procedure. In the map, water is shown in blue, urban areas are shown in light purple, vegetated farmland is light green, bare ground is orange, and forested areas are dark green.

DN	Count	Total	Percent
1	461653	461653	5.467639
2	6072863	6534516	71.924634
3	331172	6865688	3.922273
4	1298804	8164492	15.382531
5	278878	8443370	3.302923

Figure 4: Class Area Statistics for the Madison image classified using the support vector machine classification algorithm.

*Note: for both of the class area statistics tables, DN numbers of 0: unclassified, 1: urban, 2: vegetated farmland, 3: bare ground, 4: forest, and 5: water.

Discussion:

Describe what land cover classes were difficult to map or had confusion, and why this occurred. How did you solve the problem? What parameters did you manipulate during classification? Did these changes help or hurt? If you had problems in several classes, or different types of confusion, describe each and how you rectified the problem. Discuss whether you thought one method worked better than the other, and whether the post-processing steps hurt or helped in some way.

The most difficult areas to classify were the lightly vegetated areas located in the southeastern portion of the two images. These areas were difficult to classify because as a whole, they have complex spectral signatures – a combination of the vegetated farmland and bare ground classes. The confusion occurred due to the differing spectral signatures, so I attempted to create a new class that would include these lightly vegetated areas. However, after doing this I found that some other areas were now also being classified as lightly vegetated, even though they were actually bare ground or vegetated areas. In the end, I decided to just classify lightly vegetated areas as vegetated areas to avoid additional confusion between classes.

Another issue was the unclassified pixels in the maximum likelihood classification map. Originally, my training sites were individual pixels and were much too specific to help with interpolating the information to the rest of the map, resulting in a large portion of my map being unclassified. To rectify this, I chose new training sites and included larger variance in pixels. After three iterations through this process, I finally got my unclassified pixels down to .01% of the total pixels, as shown in Figure 2. I was satisfied with this and used it as my final product for this classification.

I think that the support vector machine classification worked better, as it in my comparison of my classified maps to the actual imagery, the SVM classification seems to fit the imagery much better. This could be due to the way the SVM chooses a line of maximum distance between each of the classes and fits it to the data. The post-processing steps helped to make the map more readable and accurate, as the pixels in the bare ground fields that were classified as urban were able to be reclassified to their appropriate values using a nearest neighbor comparison approach.

References

Huang, C., Davis, L. S., & Townshend, J. R. G. (2002). An assessment of support vector machines for land cover classification. International Journal of Remote Sensing, 23, 725–749.