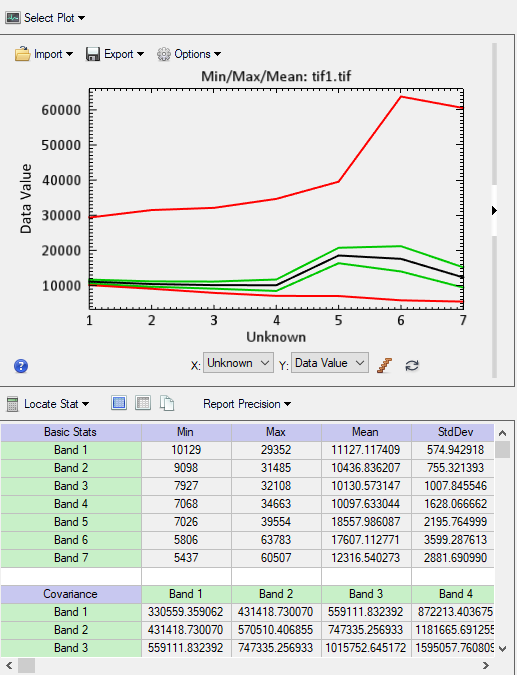
**GEOG 475 Lab 4**

Arielle Wood

**1.1**



Sort / Filter the data

Put data into ArcMap

Put data into ENVI

Classify Data

**1.2**

***1. Include your conceptual workflow (flowchart) that highlights data, information products, feature space, and classification and post classification stages. Describe the overall workflow.***

The workflow (above) includes changing the data in both arcmap and ENVI. This is necessary due to the nature of the data and the potential of both programs.

***2. Justify the computational approach for producing the information that you will use to define your feature space. How is each type of information used to differentiate between the landcover classes?***

My feature space will be defined using both BP-ANN algorithms and fuzzy C-means algorithms due to the way they interact with the data. They produce the most accurate results.

***3. Is multicollinearity a problem in your n-dimensional feature space? What evidence do you have to support your answer?***

Multicollinearity is an issue with the n-dimensional feature space. This is understood due to analysis between the slopes of the TIFF data. Many of the slopes were very similar. This is indicative of multicollinearity. This is not a major issue because the multicollinearity is minor.

***4. How does your feature space and classification results compare to the classification results from using the original spectral data as the feature space? Do you have better accuracy for some classes?***

The urban areas were, to a degree, misrepresented by many of the algorithms used. One example of this was examining classifications based upon water. In most cases, the water classes included bodies of water and the densely populated city. This is due to the nature of the data, and the nature of the algorithms used to analyze the data. The accuracy was relatively equal between my feature space and the spectral data feature space.

***5. How easy or difficult do you think it is to accurate map the environmental classes? Did your approach work?***

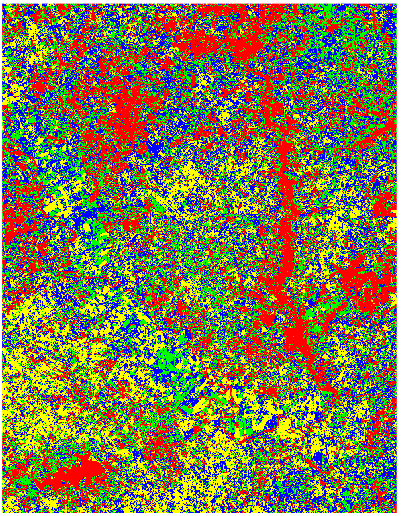
Mapping the environmental classes was, at first, easy. The algorithms did most of the work. Yet, upon further analysis, I found accurately mapping environmental classes to be difficult. This is due to the level of analysis required to maintain accuracy. In many cases, the algorithms incorrectly classified regions, and an untrained eye would potentially not have caught such mistakes. I argue that it is difficult due to the high margin for error.

***6. Can pattern recognition be used to classify environments?***

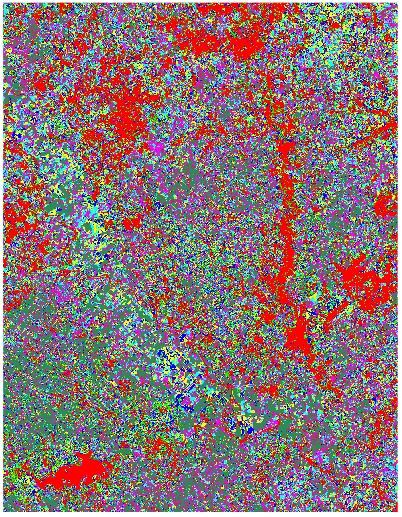
Yes, pattern recognition can be used to classify environments. It does, however, have a fundamental issue. Pattern recognition requires a test sample. The test sample is created by the user. This opens the algorithm up to human error. The maps created may have issues that began with the user selected test sample. Pattern recognition can be used for classification, but it requires a knowledgeable user to create meaningful maps.

**2.1**

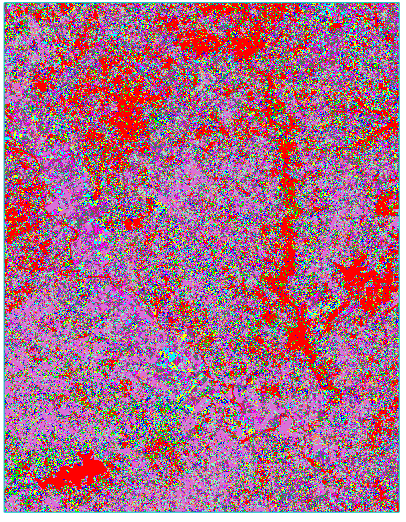
Class of 4



Class of 8

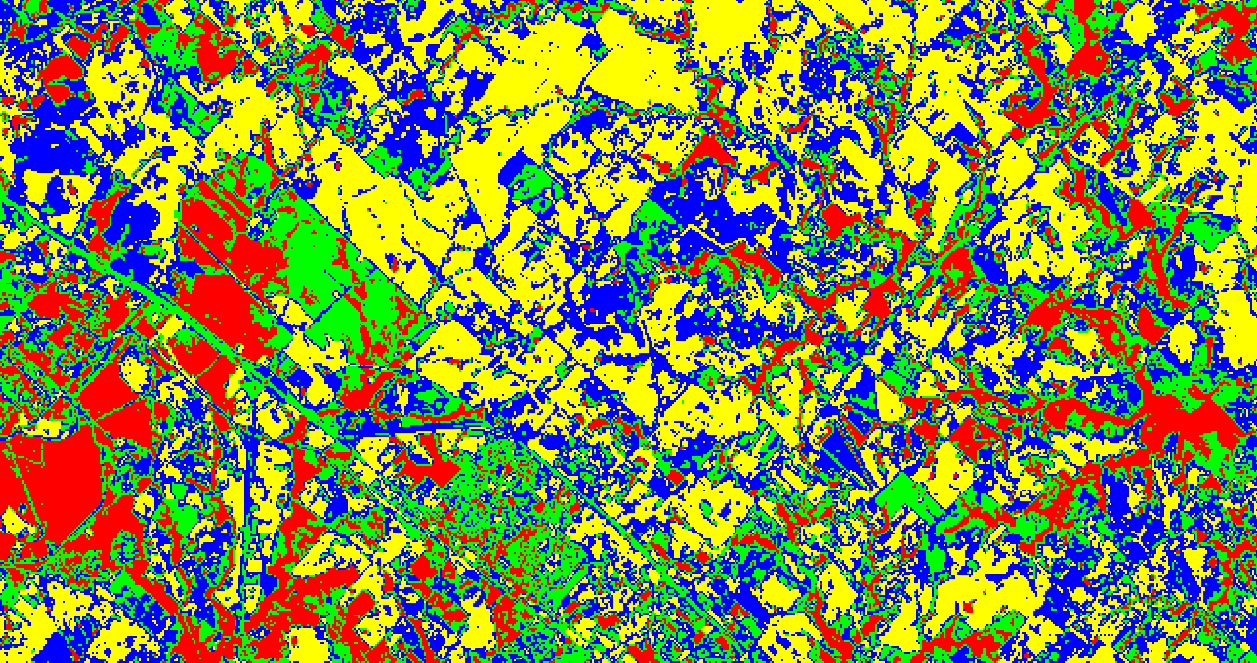


Class of 12



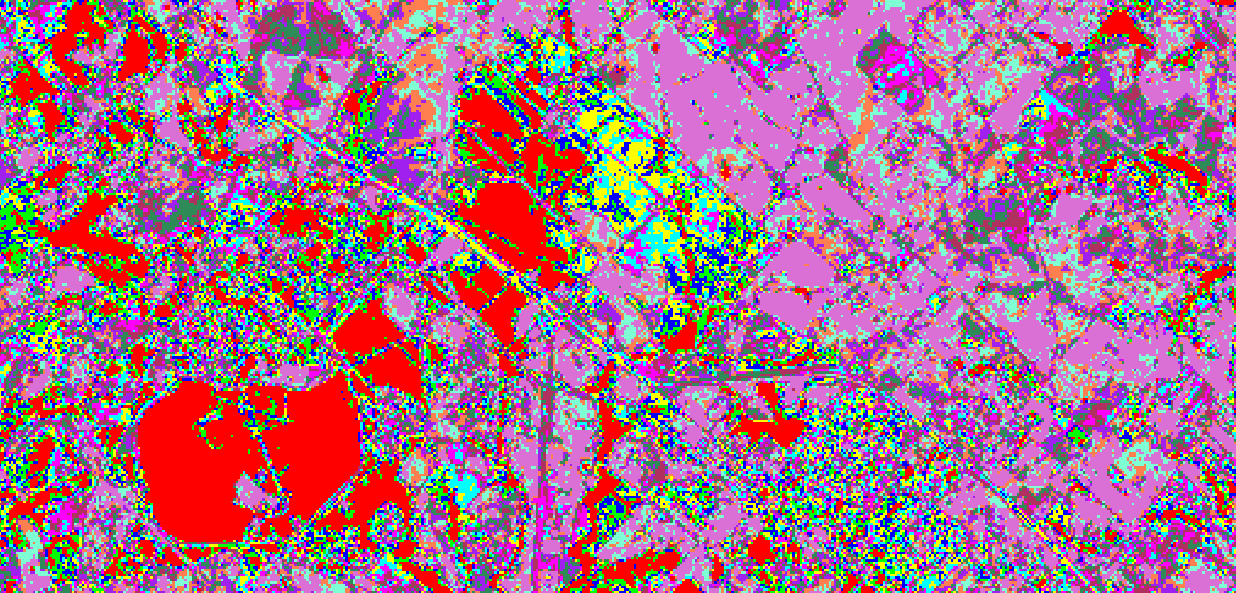
**2.2**

***1. How well does the ISODATA algorithm classify 4 real land cover classes? Do the classification patterns correspond to areas that you know represent that class? Qualitatively is the classification accuracy relative high, moderate or low?***

The isodata algorithm does not classify 4 land classes well. The classification categories must be relatively broad, thus land with both in a category would have a higher likelihood of an inaccurate representation. Land classification is more of a ‘greyscale’ than a distinct ‘black or white’ classification. One piece of land might fit in multiple categories. The smaller the number of land cover classes, the less accurate the classification. This can be distinctly noticed when zooming in on the land cover classification map. 

***2. Do the classification results improve if you 8 and 20 cluster classes. Describe what you see. Be sure to color code your classes. Can you explain what ISODATA is doing?***

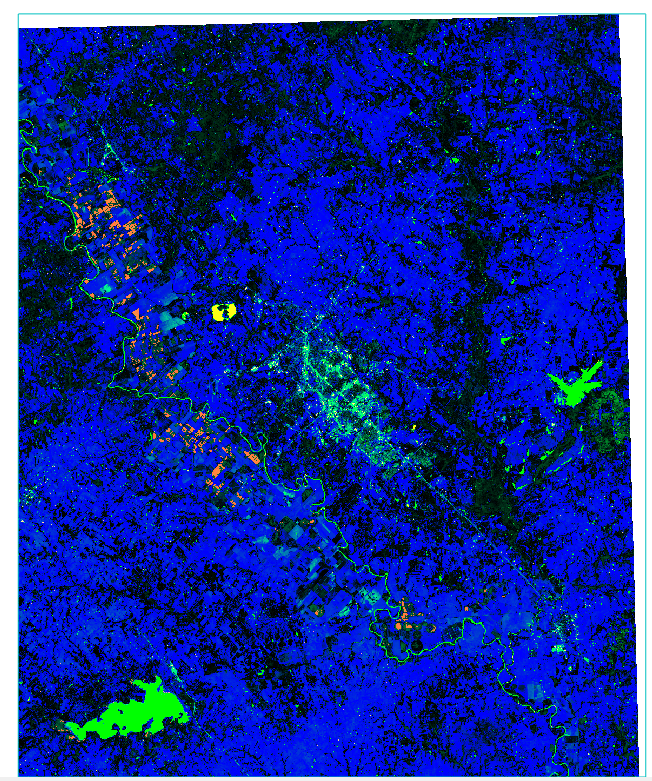
Yes, the accuracy of the classification results are directly correlated to the number of land cover classes. 8-20 land cover classes would allow for more complex classes. This means that the land would be classified closer to a gradient than as distinctly separate entities. Areas with complex land cover would not be put in the same category as areas with simple land cover. The classification would be more accurate.



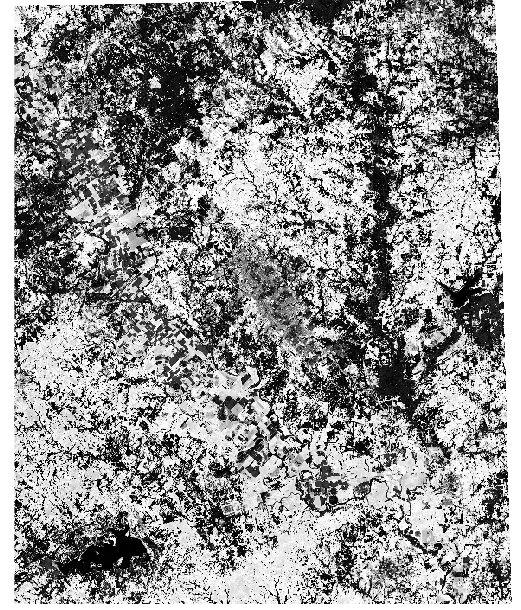
***3. What are your conclusions about using ISODATA to accurately classify land cover conditions?***

Isodata classifications are not the best or most accurate way to classify land cover conditions. Land cover is more of a gradient, than a distinct separation of land types. Land cover cannot easily be put into categories. Most land cover has some soil, some concrete, and some water. This means it is not easy not accurate. Isodata classifications would classify an area with mostly water as an area that is completely water. This is an example of why isodata classification is not the most accurate technique.

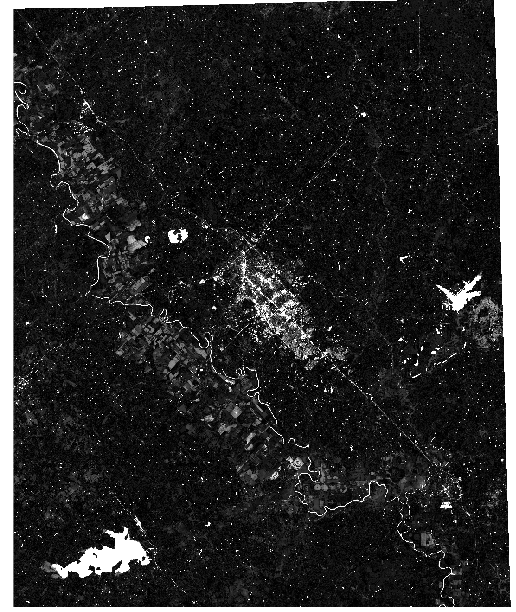
**3.1**



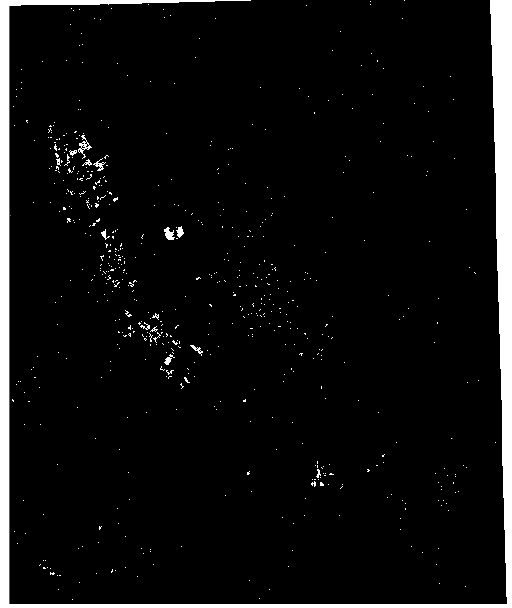
Band1



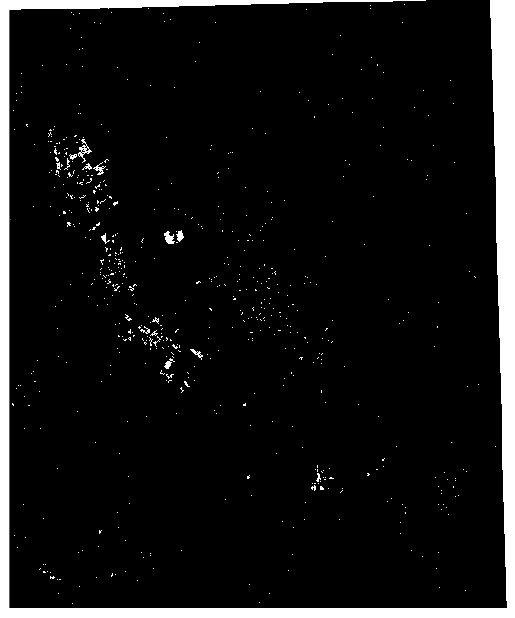
Band2



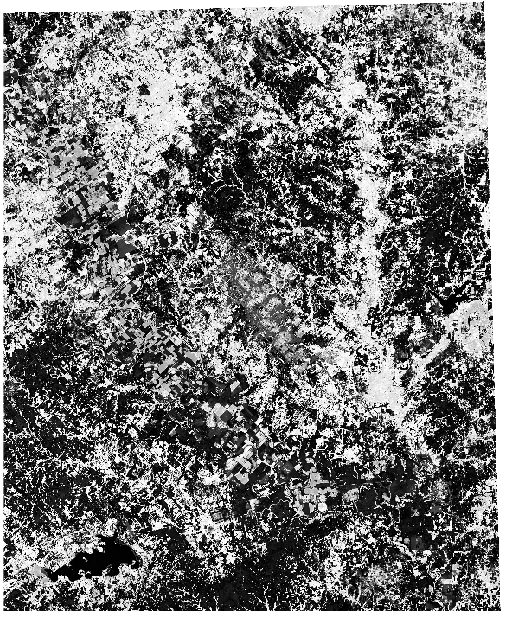
Band3



Band4



Band5



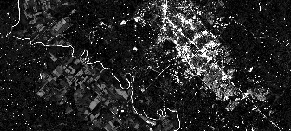
**3.2**

***1. How does the fuzzy partition matrix exponent influence the classification results?***

This algorithm is based off the brightness / reflectance of the land cover. While differing types of land cover can clearly be differentiated using color, they cannot be as easily seen using solely reflectance. Land covers giving off the same amount of reflectance / brightness may not necessarily be the same type of land coverage. This would significantly change the classification results.

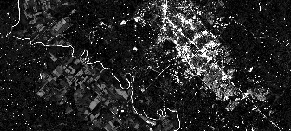
***2. How well does the fuzzy C-Means algorithm classify 4 classes? Do the classification patterns correspond to areas that you know represent that class? Qualitatively is the classification accuracy relative high, moderate or low?***

There are 5 classes within the fuzzy c-means algorithm. Class 1 is meant to target soil. Classes 2 and 3 are meant to target water coverage. Classes 4 and 5 are related to the NIR. The classification techniques for this algorithm are not the most accurate. They are based off of the brightness / reflection from the land. Areas with differing land coverage might give off the same level of reflectance throwing off the accuracy of the coverage classification. This is seen when looking at class 2. Class 2 is only meant to depict water coverage, yet it shows both a river and the city. This is due to the two types of land coverages outputting the same amount of reflectance / brightness. The classification accuracy is moderate to low.



***3. Do the classification results provide any insights into the classification accuracy? Describe the results.***

The results give significant insights into the classification accuracy. This is especially seen when analyzing the class 2 land coverage map. Class 2 is meant to target only water coverage, yet both a river and the city is clearly depicted in the visual.

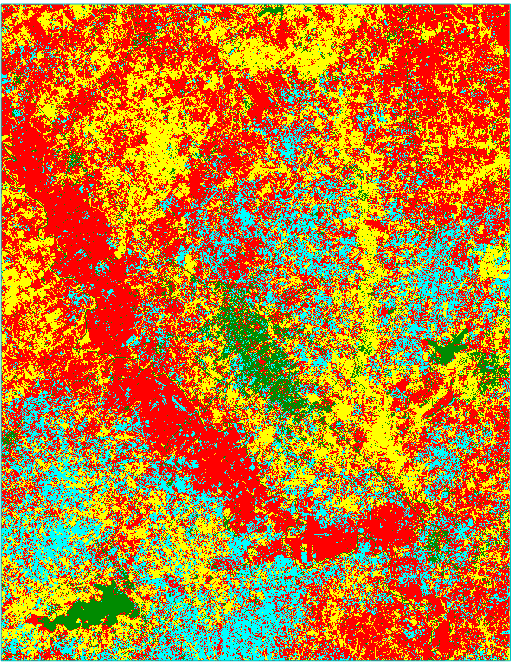


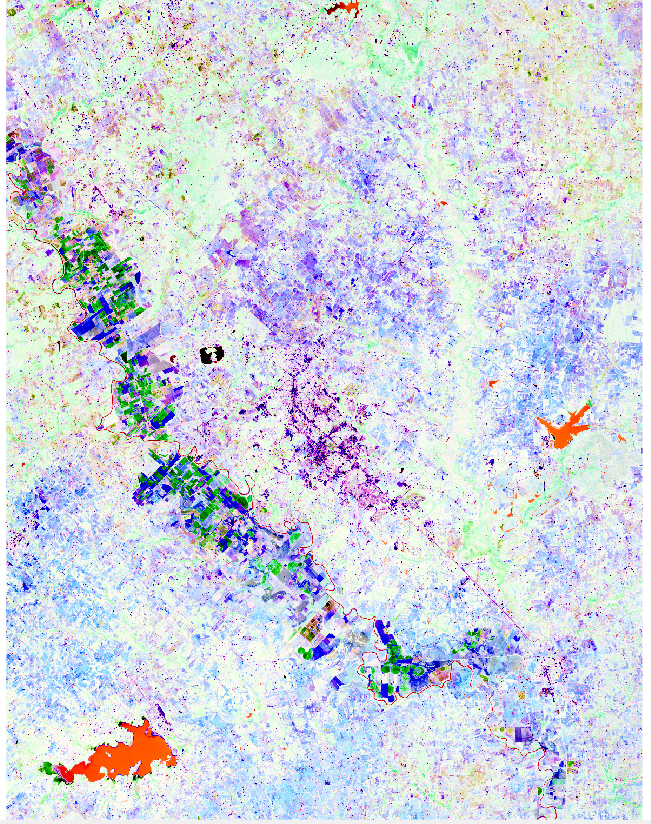
This shows that the classification accuracy cannot be very accurate. Different types of land cover can give off the same amount of brightness / reflectance. This causes an issue when trying to classify land cover using only these results. One might, mistakenly, think that the city is flooded or under water.

***4. What are your conclusions about using the fuzzy C-Means algorithm to accurately classify land cover classes?***

The Fuzzy C-Means algorithm is more accurate than the isodata algorithm, because it classifies a lot of the data in more of a gradient fashion. It is not, however, completely accurate. Different types of land cover can give off the same level of brightness / reflectance. This significantly affects the results and changes the level of accuracy.

**4.1**





**4.2**

***1. How does training sample selection influence the classification results?***

The training sample is the ROI values selected. The user tells the computer how to classify the land. This means that the user tells the computer what to classify as soil, water, concrete, and vegetation based on a small test region. The computer applies the example or test region to the entire map.

***2. How well does this algorithm classify your classes? Do the classification patterns correspond to areas that you know represent that class? Qualitatively is the classification accuracy relative high, moderate or low?***

This algorithm did a moderately accurate job of classifying the results. An error can be found in the water classification as the city is partially classified as water. The overall classification was not fully accurate, this could be due to human error or computer application error. The test regions selected were very small, and which could have been problematic. The classification patterns moderately correspond to areas I know to be certain classifications of land cover.

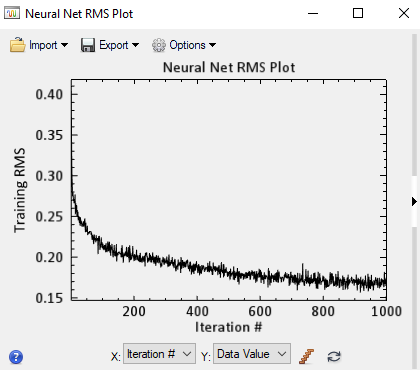
***3. Do the classification results provide any insights into the classification accuracy? Describe the results.***

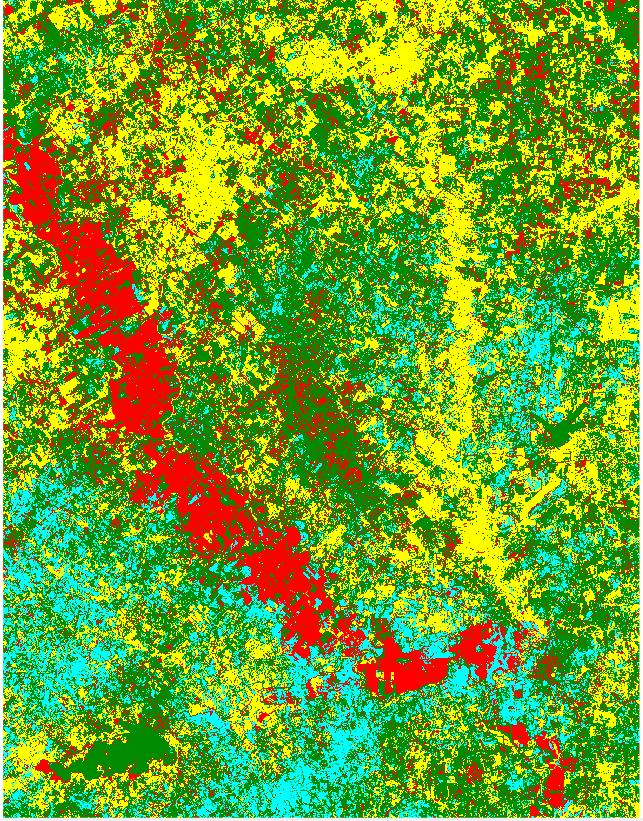
The errors in the results directly show the errors in the classification accuracy. This can be seen through looking at the water classification. Regions of the city are classified as water cover. This is indicative of a mistake in the classification of the region. It means that this classification technique is not the most accurate.

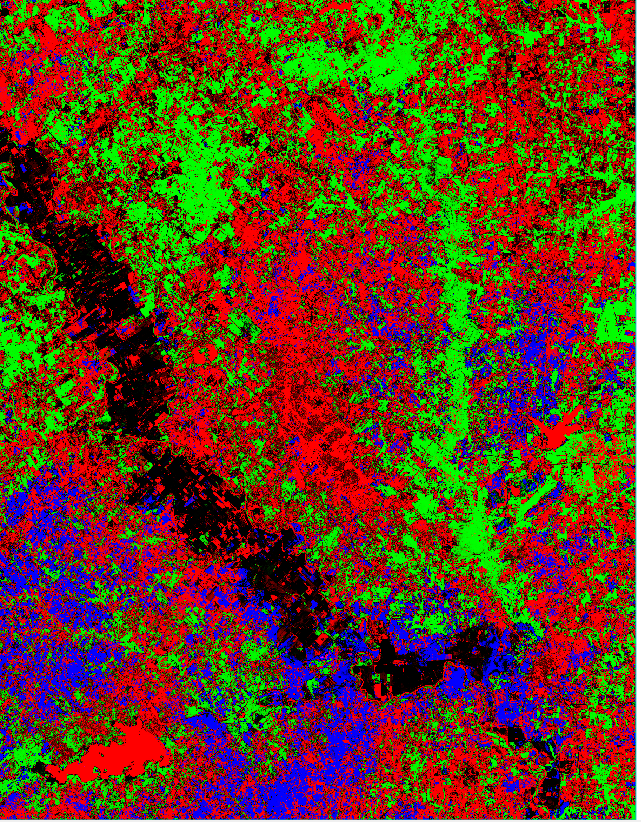
***4. What are your conclusions about using this algorithm to accurately classify land cover?***

This algorithm is moderate at classifying land. It is subject to human error which can make it a risky option for accurate results. The human component, however, could also serve to increase the accuracy. If an error was human caused, it could be fixed the same way. This makes it a moderate to good land classification technique.

**5.1**







**5.2**

***1. How does ANN structure (number of nodes in the hidden layer) affect the classification results?***

The number of nodes in the hidden layer, otherwise known as the ANN Structure affects the level of ‘learning’ done by the algorithm. The level of ‘learning’ affects the depth of the pattern recognition employed. The depth of the pattern recognition directly affects the classification results. The ANN structure is very significant to the resulting classification accuracy.

***2. How well does this algorithm classify your 4 classes? Do the classification patterns correspond to areas that you know represent that class? Qualitatively is the classification accuracy relative high, moderate or low?***

The algorithm has high accuracy. This is due to the relative or gradient nature of the algorithms way of classifying data. This is shown by how it was able to classify the 4 classes. It took into account the fuzzy membership that is integral to land cover classification. This can be noted in how some regions of the map have multiple or combining colors. The fuzzy membership is present in the map.

***3. Do the fuzzy classification results provide any insights into the classification accuracy? Describe the results.***

Yes, the fuzzy classification results show that the algorithm can handle the fuzzy membership aspect of land cover. Land classification is similar to soil classification. Changes in the classification often appear stark and distinct. This is not accurate to the landscape. Landscape changes are noted in more of a gradient. This is represented by fuzzy membership.

***4. What are your conclusions about using this algorithm to accurately classify land cover?***

This is a good classification algorithm. It can be manipulated and changed by the user via the test sample. It also takes into account the concept of fuzzy membership. This algorithm does have the potential for error residing mainly in the user component, but overall, it is a reliable algorithm to use for land cover classification.

**6.1**

***1. Compare and contrast classification results? Is there a particular algorithm that worked better than the others?***

The neural net algorithm was the most effective. It did the best job of characterizing the fuzzy membership aspect of land cover. It may also be affected by user input via the test sample. This does not introduce a large amount of error into the equation.

***2. What are your conclusions about discrete versus fuzzy classification?***

Fuzzy classification is the best classification due to the gradient nature of land cover. Discrete classification introduces too much assumption and potential error by avoiding the concept of fuzzy membership.

***3. What are your conclusions about unsupervised versus supervised classification?***

Supervised is the better classification method to pursue. It allows for user interaction thus allows for a changing and potentially higher level of accuracy. Unsupervised has a higher likelihood of being less accurate because the algorithm is not trained and has minimal user input. It is more ‘cut and dry’ and less dynamic. It has a higher likelihood of not fitting well or misrepresenting the data.

***4. If you had to do this all over again (do not do this!) describe in detail what you would do differently and what pattern recognition algorithm you would select to map environments (not landcover)? Explain the reasoning behind your answer.***

I would use the Neural Net algorithm. It did the best job in terms of characterizing the gradient or fuzzy membership aspect of the topography. It is also more adaptable due to the test samples inputted via the user. This allows for a deeper understanding and deeper level of pattern recognition. It combines the user and program into the map. It is the most accurate algorithm.