INTRODUCTION TO REMOTE SENSING

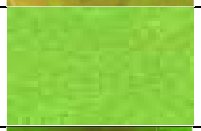
Lab 6: Supervised Classification

NAME: MST SANJIDA ALAM

a) What information classes can you identify in the color composite?

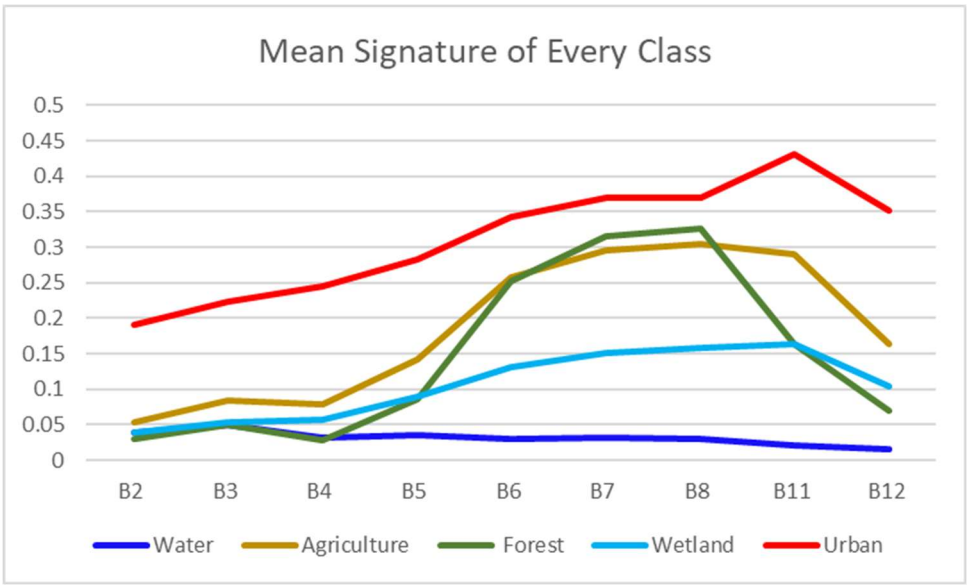
The commonly visible information classes are Vegetation, Agriculture, Water bodies, Urban Build up area and Open land. Agriculture land has yellowish green color, green represents vegetation or Forest, water appears in black color, open land or unused land area in dark yellow color; there is also some parts of urban area with a contrasting white color.

b) Create a table with the class name and ID (Label). You can use the snipping tool in Windows to capture examples of the class.

ID (Label)	Spectral class	Example
1	Water	
2	Agriculture	
3	Forest	
4	Wetland	
5	Urban	

c) Extract the mean signature for your other classes and plot them in Excel (example below)

Bands	Water	Agriculture	Forest	Wetland	Urban
B2	0.0386044	0.05424611	0.029365555	0.039526	0.190751
B3	0.04949	0.08388389	0.050181667	0.053108	0.223858
B4	0.0321661	0.07863722	0.027993333	0.056305	0.244352
B5	0.0358806	0.14128278	0.086219444	0.089288	0.283335
B6	0.0305139	0.25783	0.252976667	0.131067	0.342649
B7	0.0326883	0.29541167	0.316117778	0.151394	0.369408
B8	0.0299228	0.30495667	0.325602221	0.158218	0.370634
B11	0.0219283	0.28971	0.164466667	0.163563	0.430945
B12	0.016135	0.16469167	0.06988	0.104063	0.351549



d) what can you say about the separability of the classes?

From the starting point all the classes can be differentiated but they close to each other except urban and they begin to separate with the advances of graph and the greatest separability is visible between the bands B5 to B8. Water (Blue line) has a consistently low reflectance across all bands, and it is quite distinguishable from other classes. Agriculture (brown) and Forest (green) show similar spectral patterns, especially from B5 to B7. However, Forest has a higher mean in B6 and B7, while Agriculture shows a slight dip here. However, Wetland maintains a distinct spectral response in B2, B3, and B11, which could aid in its separation and it has lower reflectance than agriculture, forest and urban but higher reflectance than water. The Urban class, represented by the red line, shows a distinct spectral signature with a consistently high mean value across bands, peaking in B8. This strong separation, particularly in B5 to B8, makes it relatively easy to distinguish Urban areas from other classes.

e) How many training and testing points do you have?

Training = 375 and Testing = 75 points.

f) Interpret the results of the accuracy assessment.

Confusion matrix:

			Water	Agriculture	Forest	Wetland	Urban
		ID 0	ID 1	ID 2	ID 3	ID 4	ID 5
	ID 0	0	0	0	0	0	0
Water	ID 1	0	74	0	0	1	0
Agriculture	ID 2	0	0	74	1	0	0
Forest	ID 3	0	0	0	77	0	0
Wetland	ID 4	0	1	0	0	74	0
Urban	ID 5	0	0	21	0	0	49

As per confusion matrix water, agriculture, forest and wetland are correctly classified with no or minimal confusion with other classes. Most samples of urban are correctly classified but it shares few similarities with agriculture which leading to some confusions.

Training Overall Accuracy is approximately 0.933; indicating that about 93% of all the classification attempts are correct.

The Training User Accuracy represent how accurately the model classifies each land use type.

Class ID	Class name	Training Users Accuracy	Training User Accuracy (%)
1	Water	0.9867	98.67%
2	Agriculture	0.7789	77.89%
3	Forest	0.9872	98.72%
4	Wetland	0.9867	98.67%
5	Urban	1	100%

High Training Users Accuracy shows across all the classes except agriculture due to crop variability or mixed land use pattern. But overall, most classes have high accuracy and close to 1.0 or 100% which indicate that the model effectively identifies all these land use types; and reflects correct model performance.

Training Producers accuracy is 98% to 100% (0.9867 to 1) for most of the classes (water, agriculture, forest, wetland) except urban (0.7 or 70%). Overall, every class shows good training producers' accuracy which indicates the model is performing well across all the classes.

g) Transform the error matrix from the console to a table (using e.g. Excel)

Validation Error Matrix MD

			Water	Agriculture	Forest	Wetland	Urban
		ID 0	ID 1	ID 2	ID 3	ID 4	ID 5
	ID 0	0	0	0	0	0	0
Water	ID 1	0	15	0	0	0	0
Agriculture	ID 2	0	0	15	0	0	0
Forest	ID 3	0	0	0	13	0	0
Wetland	ID 4	0	0	0	0	15	0
Urban	ID 5	0	0	5	0	0	15

Validation Overall Accuracy MD: 0.935897435897435

H) Repeat the above changing the code to evaluate the results of the Random Forest classification.

Confusion matrix:

			Water	Agriculture	Forest	Wetland	Urban
		ID 0	ID 1	ID 2	ID 3	ID 4	ID 5
	ID 0	0	0	0	0	0	0
Water	ID 1	0	75	0	0	0	0
Agriculture	ID 2	0	0	75	0	0	0
Forest	ID 3	0	0	0	77	0	0
Wetland	ID 4	0	0	0	0	75	0
Urban	ID 5	0	0	1	0	0	69

All classes are correctly classified. Only urban class shares very minimal similarity with agriculture. The model exhibits high accuracy across Water, Agriculture, Forest, Wetland, and Urban classes, indicating that it effectively distinguishes these classes.

Training Overall Accuracy is 0.9973 which indicates that 99.73% is accurate. It shows excellent accuracy overall and performs very well on the training data. Training Users Accuracy and Training Producer's Accuracy (RF) value of almost all the class are near to 100% except agriculture (98.6% in Users accuracy) and urban(98.5% in producers accuracy) which indicating perfect classification across all the classes.

Validation Error Matrix RF

			Water	Agriculture	Forest	Flooded Forest	Wetland
		ID 0	ID 1	ID 2	ID 3	ID 4	ID 5
	ID 0	0	0	0	0	0	0
Water	ID 1	0	15	0	0	0	0
Agriculture	ID 2	0	0	15	0	0	0
Forest	ID 3	0	0	0	13	0	0
Wetland	ID 4	0	0	0	0	15	0
Urban	ID 5	0	0	2	0	0	18

Water, Agriculture, forest and wetland class most of the samples are classified well but urban shares minimal similarity with agriculture and very minor adjustment of training site selection can make it perfect. Besides, Validation Overall Accuracy RF is 0.9743589743589743 which means 97.74% of this classification is performing very well and accurate.

i) What are the major differences between the two classifications?

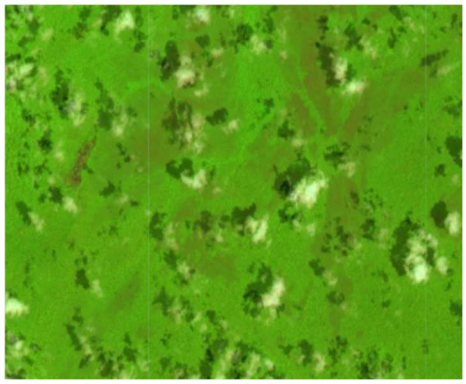
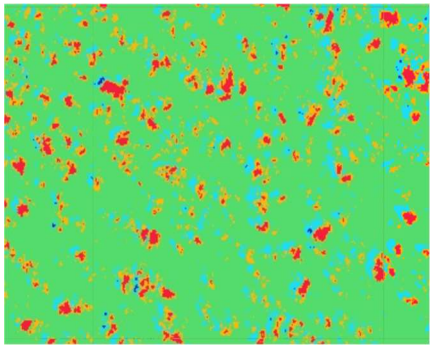
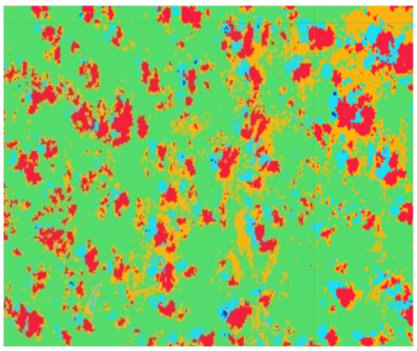
Two classification Minimum Distance (MD) and Random Forest (RF) is two different algorithm bases. MD is a simple machine learning approach based on Euclidean distance. It efficient for small datasets and may not capture complex patterns. MD classification rely on nearest neighbor similarities which simplifies the overall result.

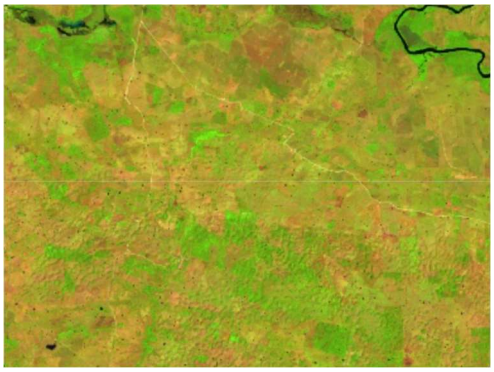
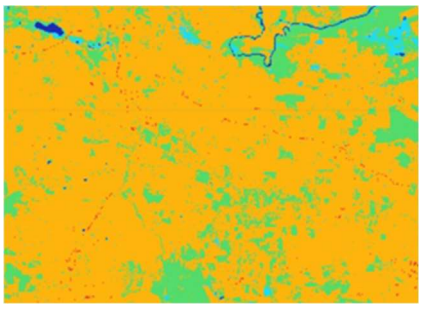
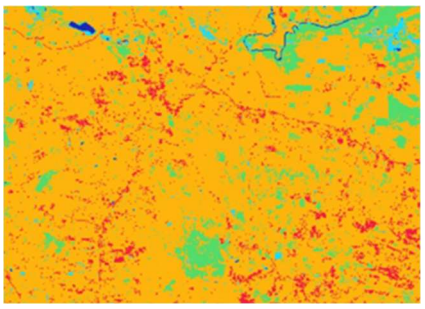
RF can build multiple decision trees during training and it is a non-parametric method that can handle non-linear relationships and very effective for complex boundaries and land use classification. That's why more distinct boundaries are found in RF classification which is missing in MD classification. On the other hand, RF value is more accurate (97.74%) than MD.

j) Can you find areas in the image that show these differences?

Pictures of area of image that shows these differences are following:

Masked False Color Minimum Distance Random Forest

Masked False Color Minimum Distance Random Forest

k) Visually inspect the color composite in those areas, which algorithm do you think is correct for those examples?

Visually inspecting the color composite in those areas, I think the Random Forest algorithm is correct for those examples.

Random Forest, its color composite looks more detailed compared to the Minimum Distance which provide the sense of the actual land cover of the area. Hence, my idea that Random Forest classification is a better option in this type of land cover.

l) Discuss the importance of selecting appropriate training data for supervised classification and explain the criteria for selecting training samples. This explanation can reflect on your training sample.

The importance of selecting appropriate training data for supervised classification is crucial to make accurate, generalized, efficient and unbiased model. Beside selecting appropriate training data is highly required for covering the full range of classes, & variations, reducing computational load and redundancy.

Criteria for Selecting Training samples (based on my study): I maintain the following criteria for selecting training sample;

1. Select Relevance sample that ensures diversity of class and proper presentation of the area
2. Collect samples from whole of the study area
3. Choose Sufficient sample size to avoid overfitting or underfitting
4. Select equal number samples from every class, so that all class can be represented evenly and equally.

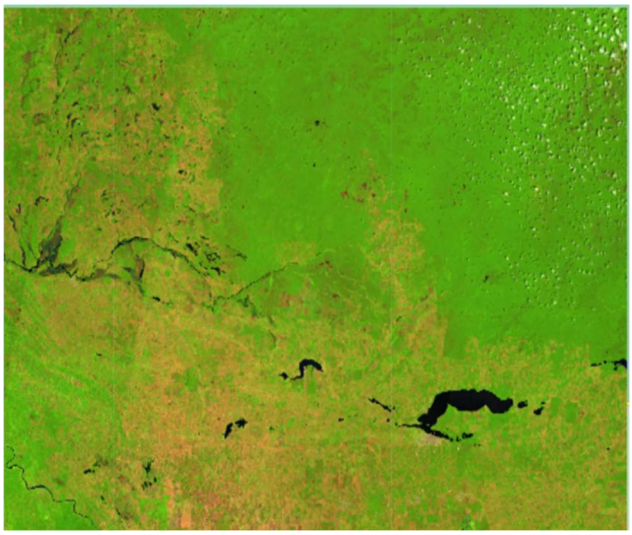
m) What is the role of spectral bands in supervised classification? How does band selection impact classification accuracy? You can try removing some bands and compare the results.

Each spectral band captures different specific information which highlight various physical properties and land cover types. That's why, Spectral bands are crucial in supervised classification as they provide unique information about the features being classified, and enhancing the separability between different classes.

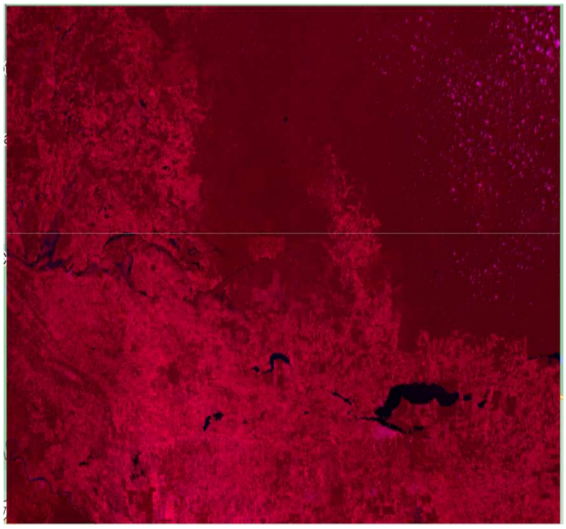
The selection of spectral bands has a direct impact on classification accuracy of the model. Appropriate band selection can improve classification accuracy by including relevant information and excluding redundant or noisy data. On the other hand, using irrelevant bands can increase computational complexity and reduce accuracy.

By removing some bands, the following difference is found:

Bands 11, 8 and 4



Band 11 and 4



Eliminating certain bands and comparing outcomes help to identify which wavelengths are useful to ensure class differentiation. By using proper bands, classification models can become more accurate, efficient, and resilient.

n) Explore the impact of changing the number of trees in the random forest algorithm on the classification accuracy.

Modifying the number of trees in a Random Forest algorithm significantly affects classification accuracy. Generally, as the number of trees increases, the model tends to improve in accuracy due to its ability to reduce overfitting and better generalize across data.

However, beyond a certain point, increasing the number of trees yields diminishing returns. This is because Random Forests are ensemble models that tend to converge with a sufficiently large number of trees, meaning further additions may not significantly impact accuracy. On the other hand, using too few trees can lead to underfitting, as the model may not capture enough patterns from the data, resulting in lower classification accuracy. Choosing an optimal number of trees is crucial to balancing high accuracy with computational efficiency, ensuring that the Random Forest model performs at its best.

Tree numbers change from 300 to 2.

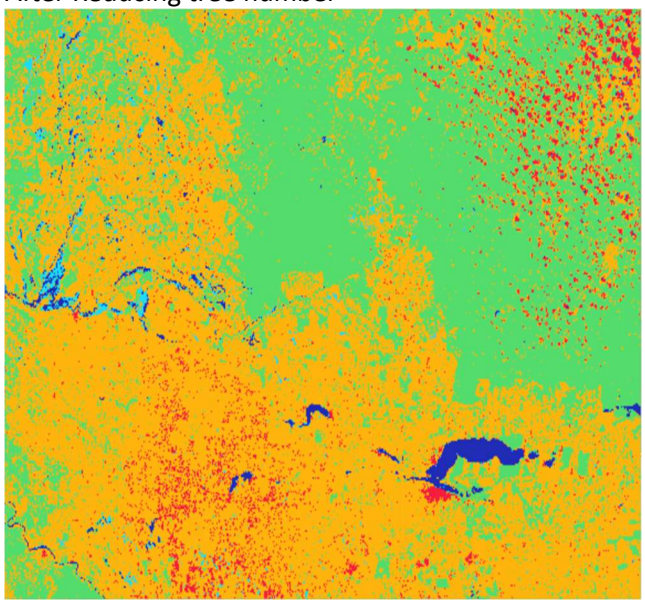
At first sight there is not much change in the composite when modifying the number of trees in the Random Forest algorithm but if zoom into the image, it can perceive very subtle changes in the extension of some classes.

```
158 var classifierRF = ee.Classifier.smileRandomForest(2,2).train({
159   features: training,
160   classProperty: 'Label',
161   inputProperties: bands
162 });
```

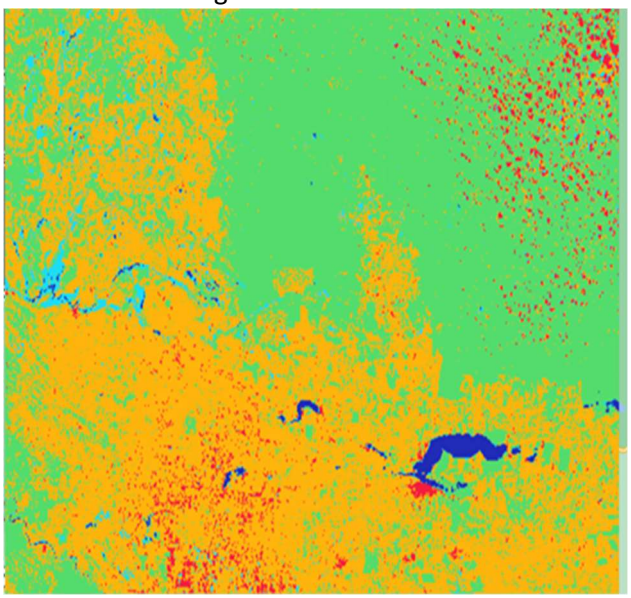
```
Validation Overall Accuracy RF:
0.9615384615384616

Validation Users Accuracy RF:
```

After Reducing tree number

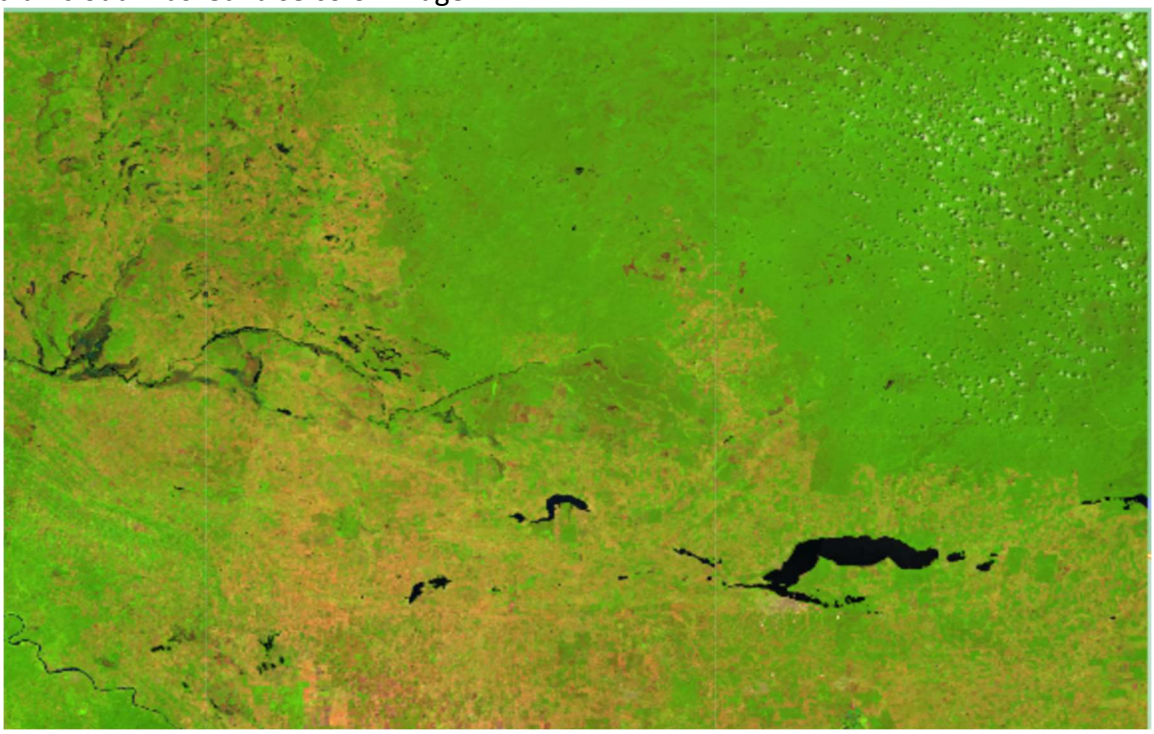


Before reducing Tree Number

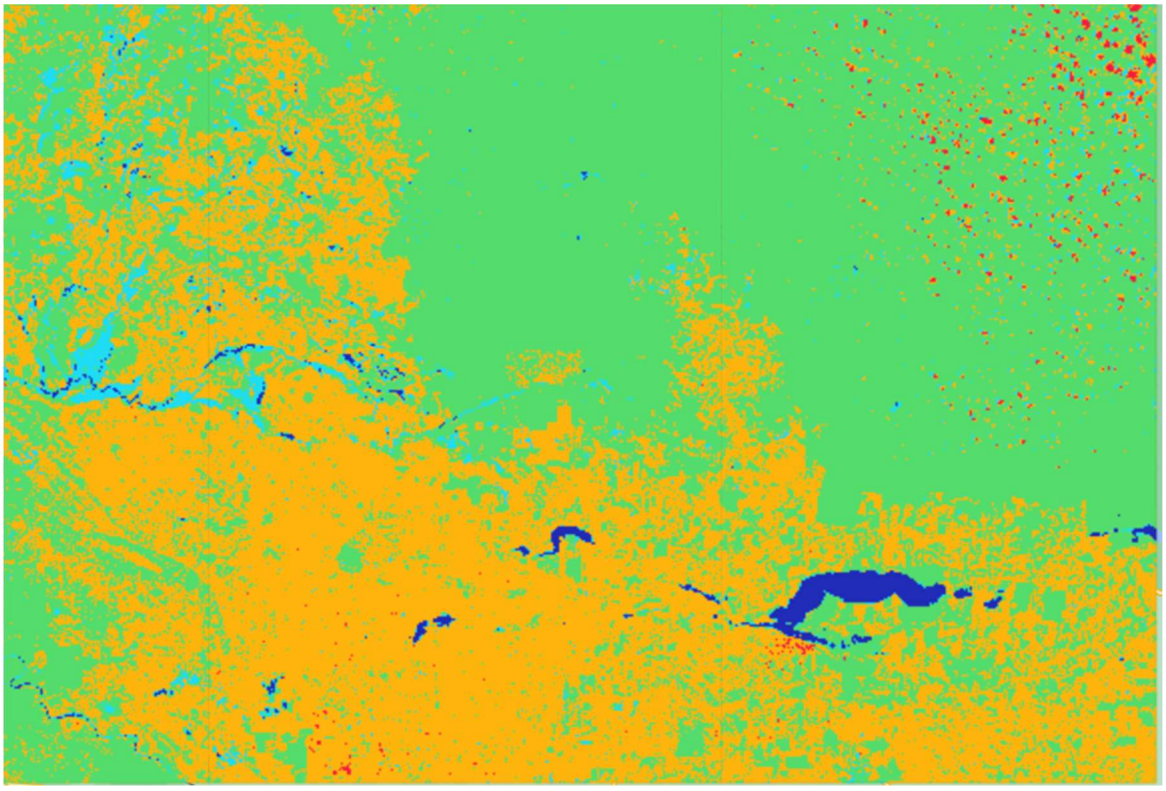


MAP OUTPUT

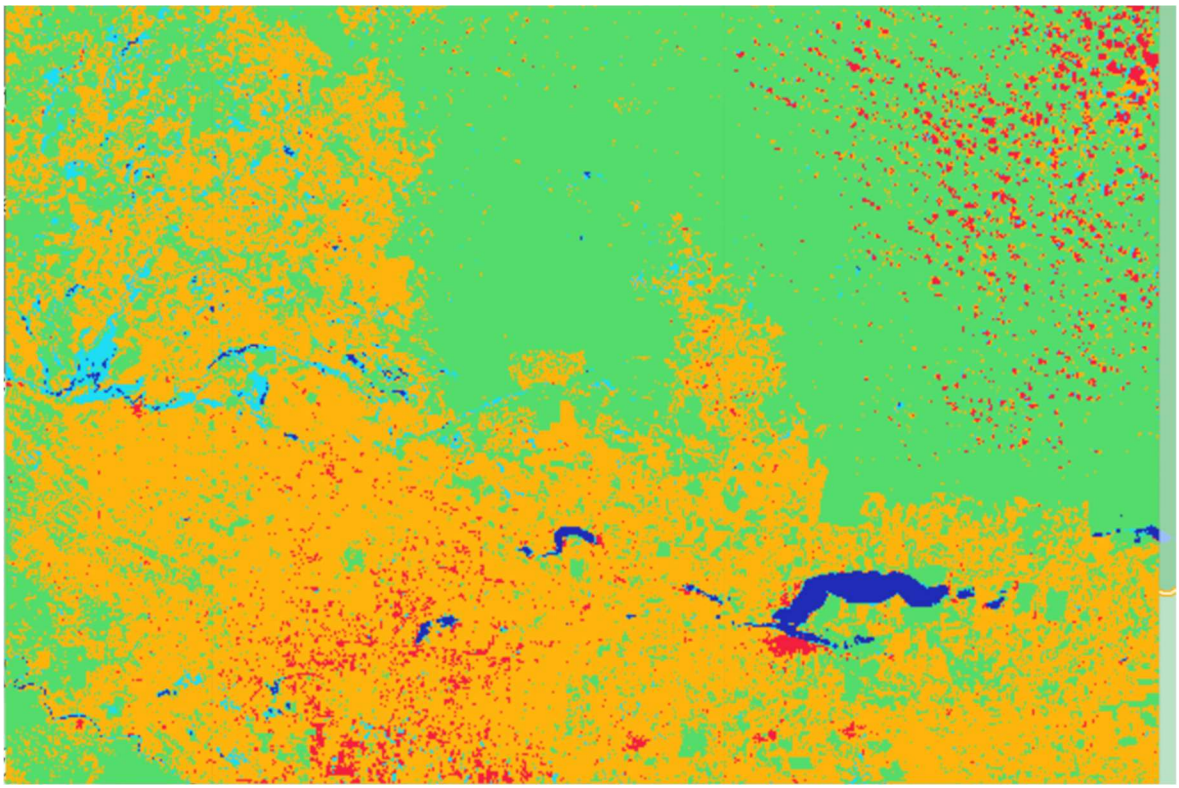
Median cloud masked false color image



Land Classification: Minimum Distance



Land Classification: Random Forest



Link: https://code.earthengine.google.com/be530803c975bc6c8e1fbe05c15b1d84