Jet Propulsion Laboratory
California Institute of Technology

Mapping Open Water Bodies with Optical Remote Sensing

M. O'Donnell [1,2] and E. Podest [1]

1.Jet Propulsion Laboratory, California Institute of Technology   2 Alliance Gertz-Ressler High School, Los Angeles, CA

Background Information

Background

- Both dengue fever and malaria are caused by a bite from an infected mosquito.
- Every year between fifty to one hundred million people are affected by dengue fever.
- People who live in tropical climates are especially at risk.

Anopheles

Why should I care?

- The mosquitos that cause dengue fever have been recently found in California, Texas and Florida, so the populations of these states could be potentially at risk for dengue. (NASA 2014).

What can be done?

- Predicting the risk of mosquito caused disease outbreaks is a required step towards their control and eradication.
- The presence of mosquitos directly correlates to occurrence of mosquito borne disease.

How can we predict the presence of mosquitos?

- The coverage and persistence of open water is one of the primary indicators of conditions suitable for mosquito breeding habitats.

- A way to find open water bodies and classify them easily and consistently, based on their likelihood of being a mosquito breeding ground, is needed.

Research idea and questions

- Remote sensing may be a way to find and classify open water bodies.
- Can open water bodies be mapped with remote sensing? Can this mapping help predict where mosquitos breed?

Previous Research

A look at what has been done on this topic

Land Cover Mapping Has Been Successfully Done using Random Forest Software

- Successful land cover classification using Landsat data was successfully done by using Random Forest (RF) in R (Hayes, M., Miller, S. and M. Murphy, M. (2014).

- Satellite imagery also using RF in Fortran was used to produce a thematic map of wetlands throughout Alaska. (Whitcomb, J., Moghaddam, M., McDonald, K., Kellndorfer, K.J., and Podest, E (2007)

Open Water Body Mapping

- The coverage and persistence of open water is currently a poorly measured variable due to its spatial and temporal variability across landscapes, especially in remote areas.

Our Research

Research Plan

- To develop a methodology in R for classifying open water bodies using a decision tree approach on Landsat images.

- The Random Forest decision tree algorithm in R will be used.

Data from Satellite Observations

- Satellite images from Landsat 7 can provide needed data for potential classification of open water bodies and their color properties.

Landsat 7	Wavelength (micrometers)	Resolution (meters)
Band 1	0.45-0.52	30
Band 2	0.52-0.60	30
Band 3	0.63-0.69	30
Band 4	0.77-0.90	30
Band 5	1.55-1.75	30
Band 6	10.40-12.50	60 * (30)
Band 7	2.09-2.35	30
Band 8	.52-.90	15
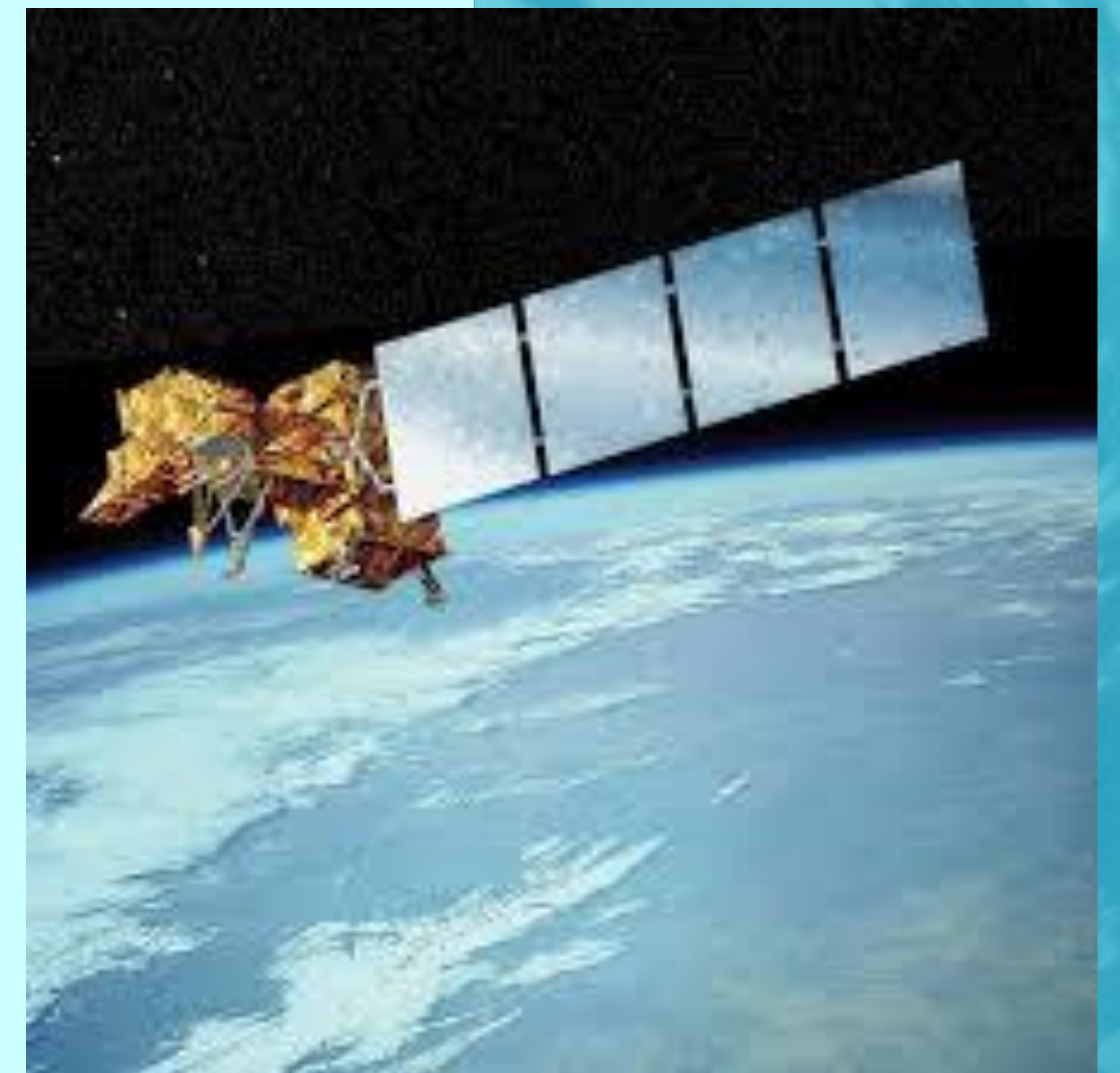
Landsat 7 Image File

Actual size of image 45 km x 45 km

The above image is an RGB subset (of bands 3,2,1) from a Landsat 7 image over Ethiopia. This area was the focus of our study.

Use of Random Forest Algorithm/Software

- Random Forest (RF) consists of a collection or ensemble of simple tree predictors, each capable of producing a response when presented with a set of predictor values. For classification problems, this response takes the form of a class membership, which associates, or classifies, a set of independent predictor values with one of the categories present in the dependent variable.
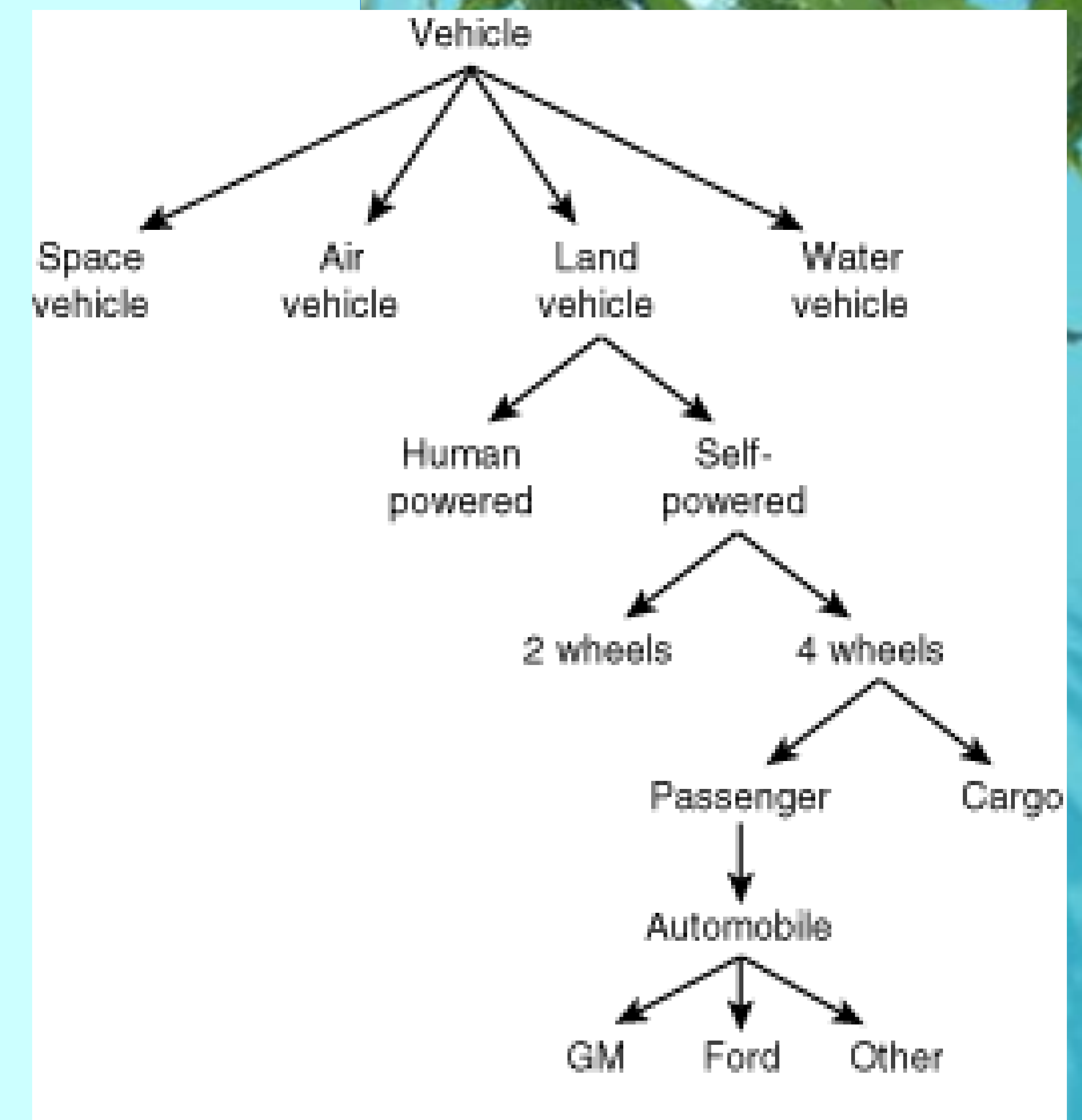(Breiman, L, Cutler, A. (2004)

- RF will be used to classify open water bodies and their color properties from Landsat images using a decision tree classification approach.

- We will use a supervised classification approach so a truth table is needed to train the classifier.

- ENVI was used to derive the truth table

Decision Trees in RF

Simple decision tree example

- A decision tree is an algorithm for categorizing.

- RF's decision tree algorithm makes a "forest" of decision trees from data based on the truth table.

  o Each pixel in the image to be classified is run through the forest. Each tree in the forest assigns the pixel a class and the class that occurs the most times over all trees in the forest is assigned to that pixel.

  o The classification error rate is low for this algorithm.

Use of ENVI Software

- ENVI is an image visualization and analysis software that in this case is used to the generate the training areas (truth) to be input into Random Forest. (Exelis, 2015)

- We visually identified the truth areas. For example, we selected areas representing water bodies within the image and in addition we identified them according to their color properties.
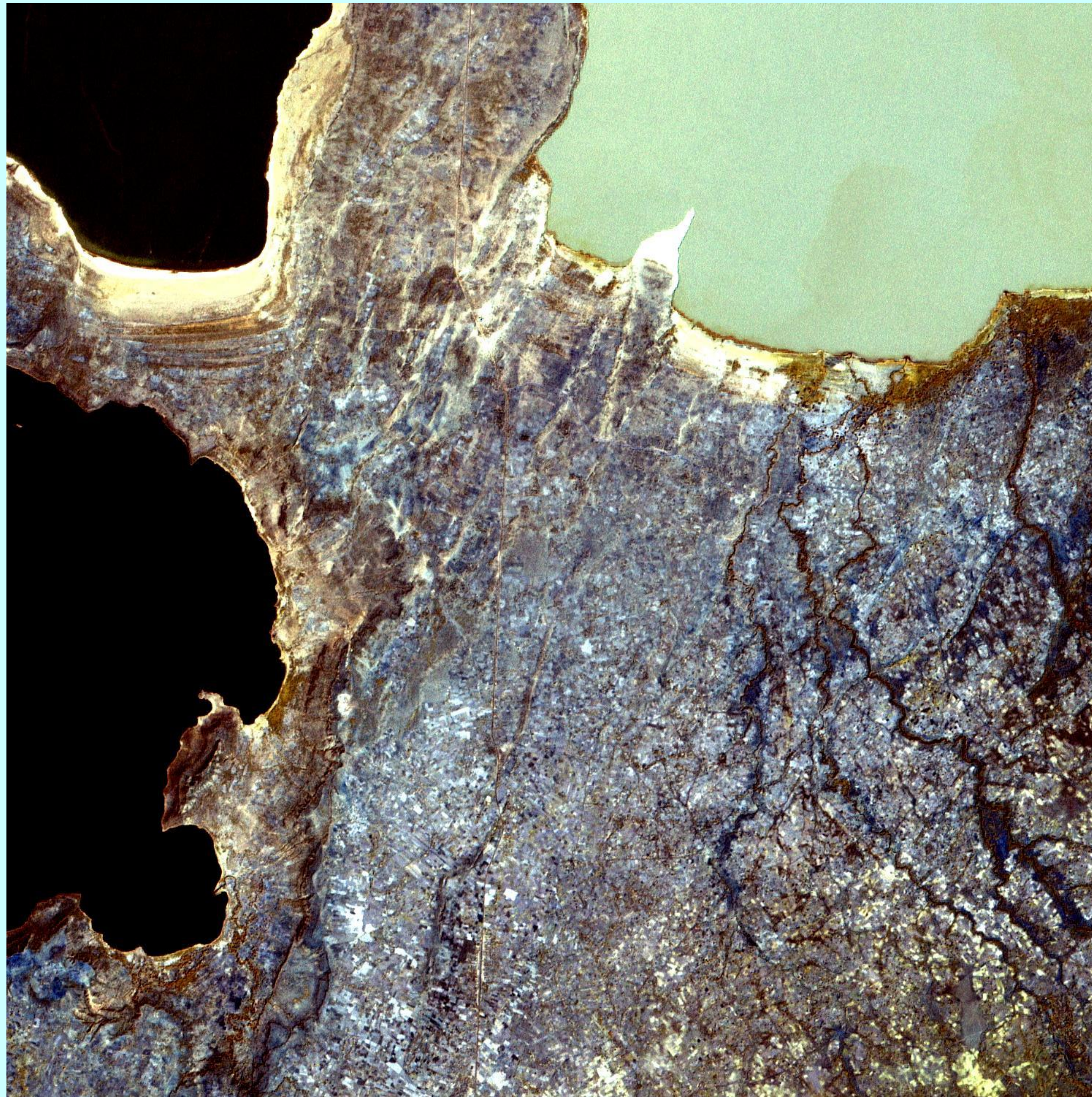
R Code

- Code was developed in this project using the R language.

- R is open source software specifically designed for use in statistics applications (The R Project, 2015)

R Code Development

- R Code was developed to:
  - Input Landsat 7 tif image file into the R environment and convert it into an R data frame
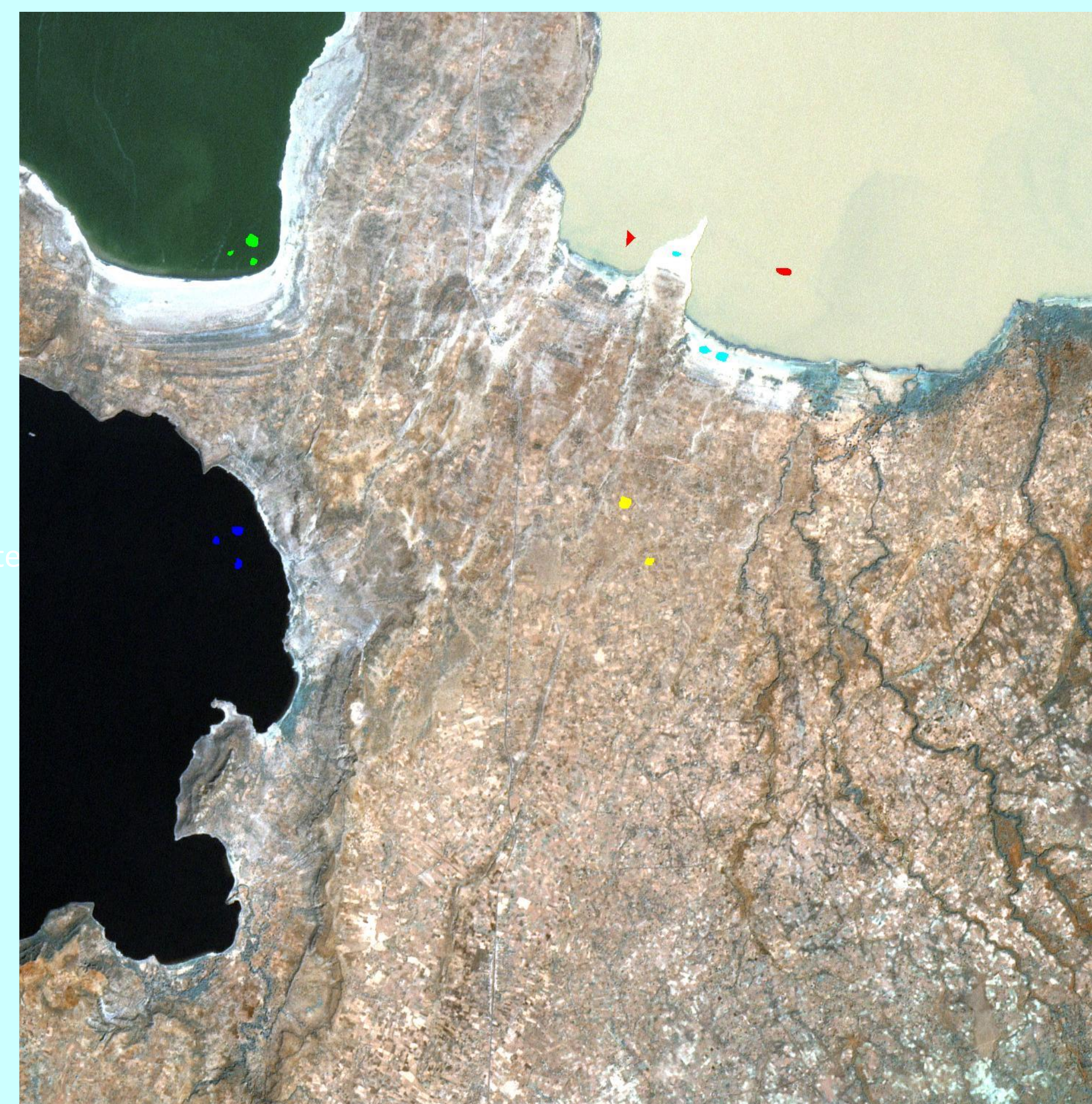

Landsat 7 image that was input into our R code
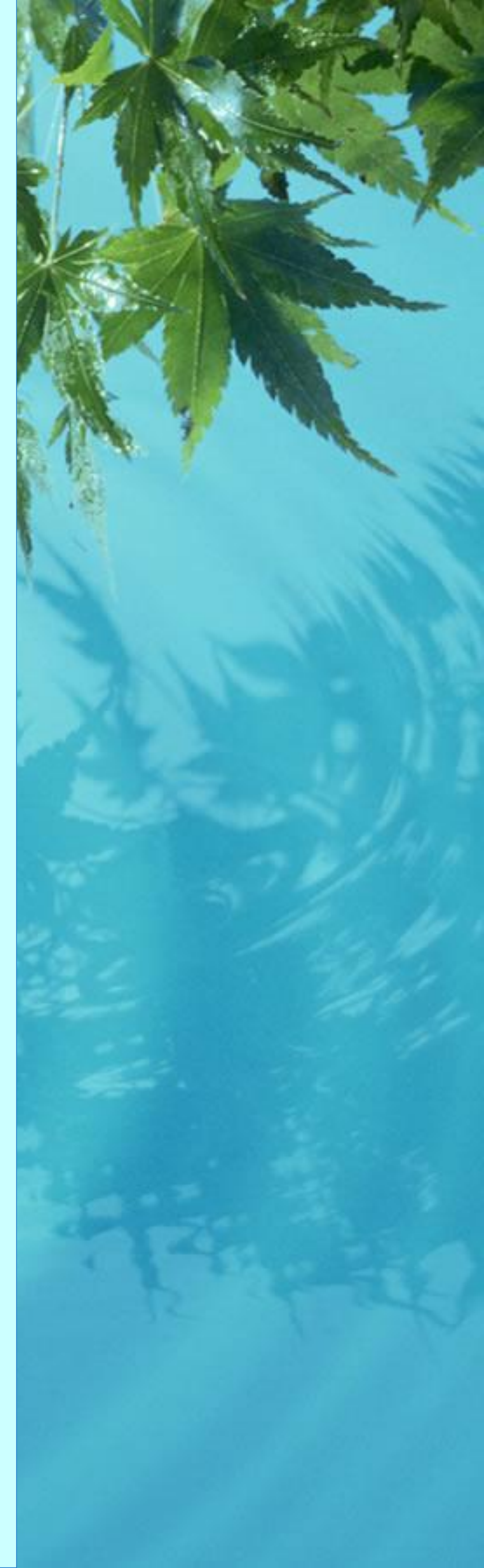
R Code Development (continued)

o Input the truth table created using ENVI and convert it into an R data frame

Landsat 7 image

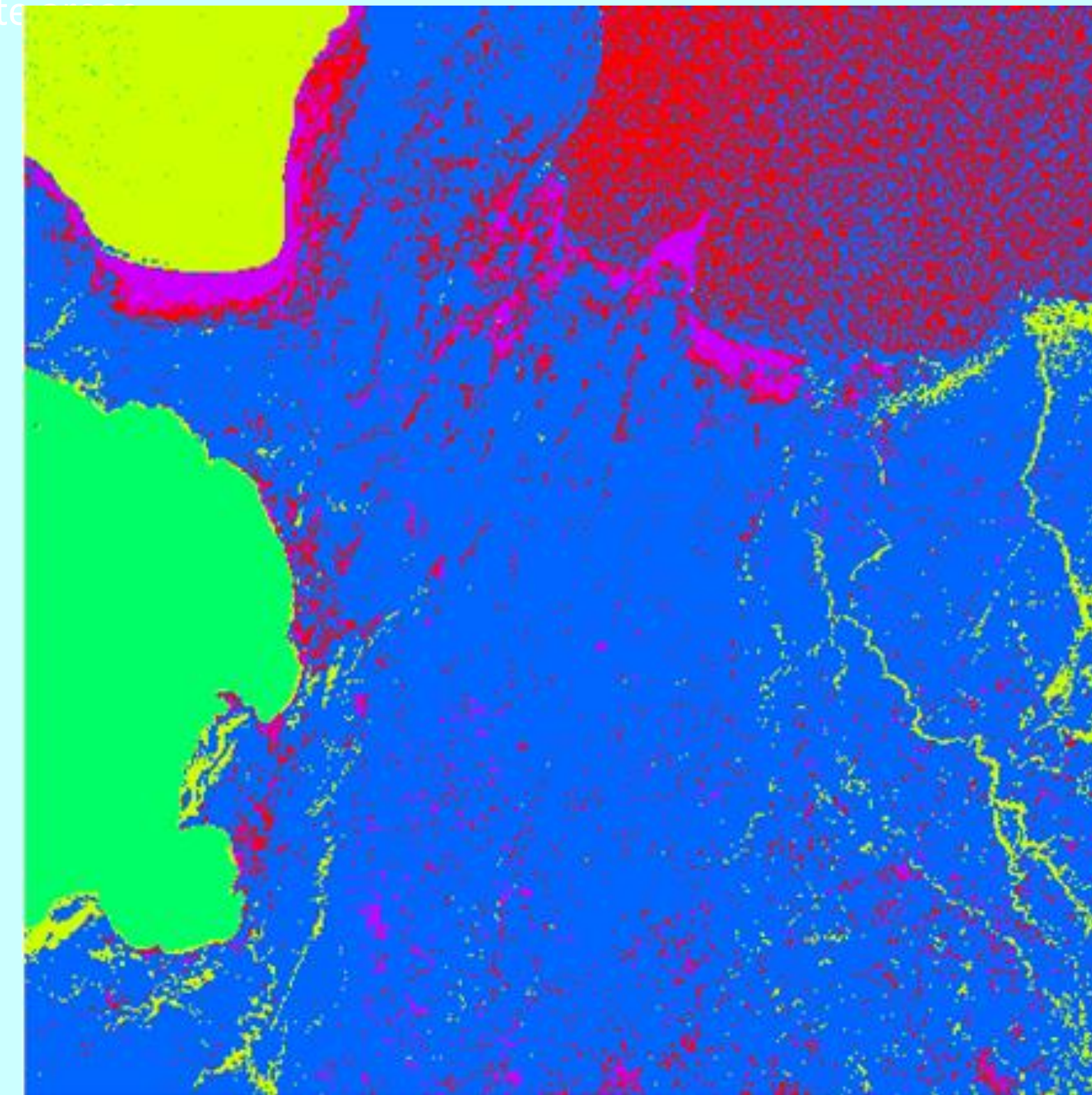Landsat 7 image with the training areas delimited

R Code (Continued)

o Run RF to create a decision tree forest

o Use the decision tree forest created to classify the Landsat image

o Create a colored pixelated map of the classified Landsat image
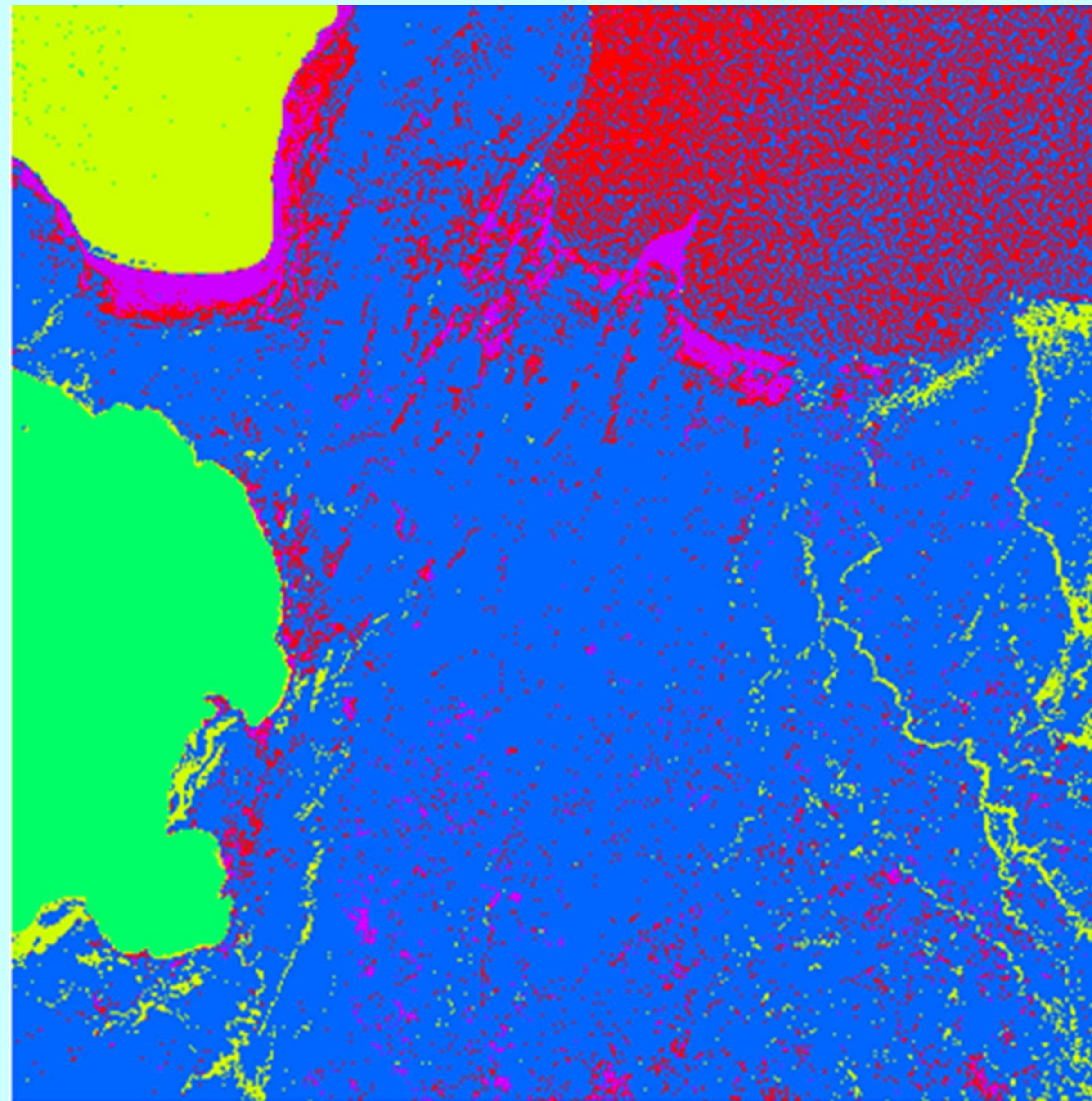

Input file (from previous slide) –RGB Bands 3,2,1 from Landsat 7

Output File – Pixel map with classifications (colors) generated from our 'R' code

R Output: Key to Mapping of Open Water Bodies

COLOR	REPRESENTS
Red	Brown water
Yellow	Green water
Green	Black water
Blue	Land
Magenta	Bare soil

Next Steps

To Do

- Analyze, test and refine the R code.

- Use the developed R software to input additional data into RF and analyze this output for trends to see if open water bodies can be mapped correctly using this technique.

References

References

- NASA DEVELOP National Program (Producer). (2014, April 2). .Brazil Health and Air Quality. Video retrieved from https://www.youtube.com/watch?v=BuKB7o7zD-4

- Matthew M. Hayes, Scott N. Miller & Melanie A. Murphy (2014) "High-resolution Landover Classification using Random Forest",   *Remote Sensing Letters*, 5:2, 112-121 http://dx.doi.org/10.1080/2150704X.2014.882526

- Whitcomb, J., Moghaddam, M., McDonald, K., Kellndorfer, K.J., and Podest, E (2007) "Wetlands Map of Alaska Using L-Band Radar Satellite Imagery " http://www.tandfonline.com/loi/trsl20

- Exelis Software (2015), ENVI Available from http://www.exelisvis.com/ProductsServices/ENVIProducts/ENVI.aspx

- Breiman, L, Cutler, A. (2004). Random Forests (Version 5.1) [Software]. Available from http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm

- The R Project for Statistical Computing (2015). Available from http://www.r-project.org/