Forest Structure and Bird Nesting Habitat Derived from LiDAR Data

Doug Newcomb USFWS, Dr. William Hargrove at the USDA Eastern Forest Threat Center, Forrest Hoffman, and Dr. Jitendra Kumar with Oak Ridge National Laboratory

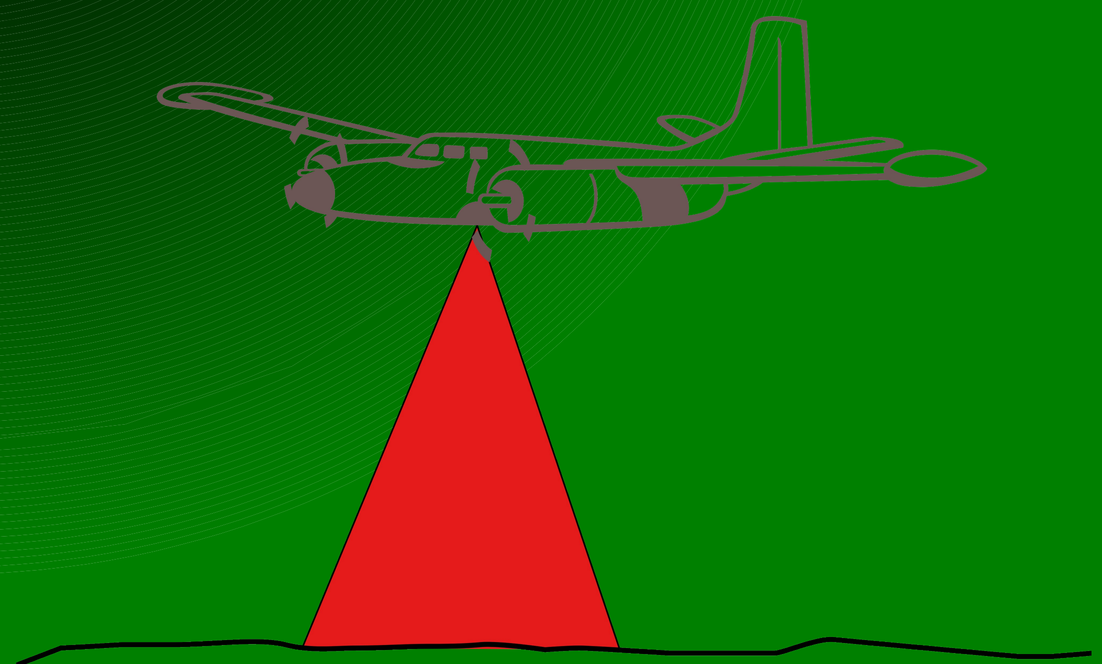
NC Floodplain Mapping Project

Used LiDAR Technology from 2001 to 2006 to map the ground surface elevation

What is LiDAR?

LiDAR devices are generally mounted in airplanes and data is collected as the airplane flies across a landscape in lines that overlap the scanned areas

Distance to Ground =(Velocity (Speed of light through air) / Time from pulse to return) /2



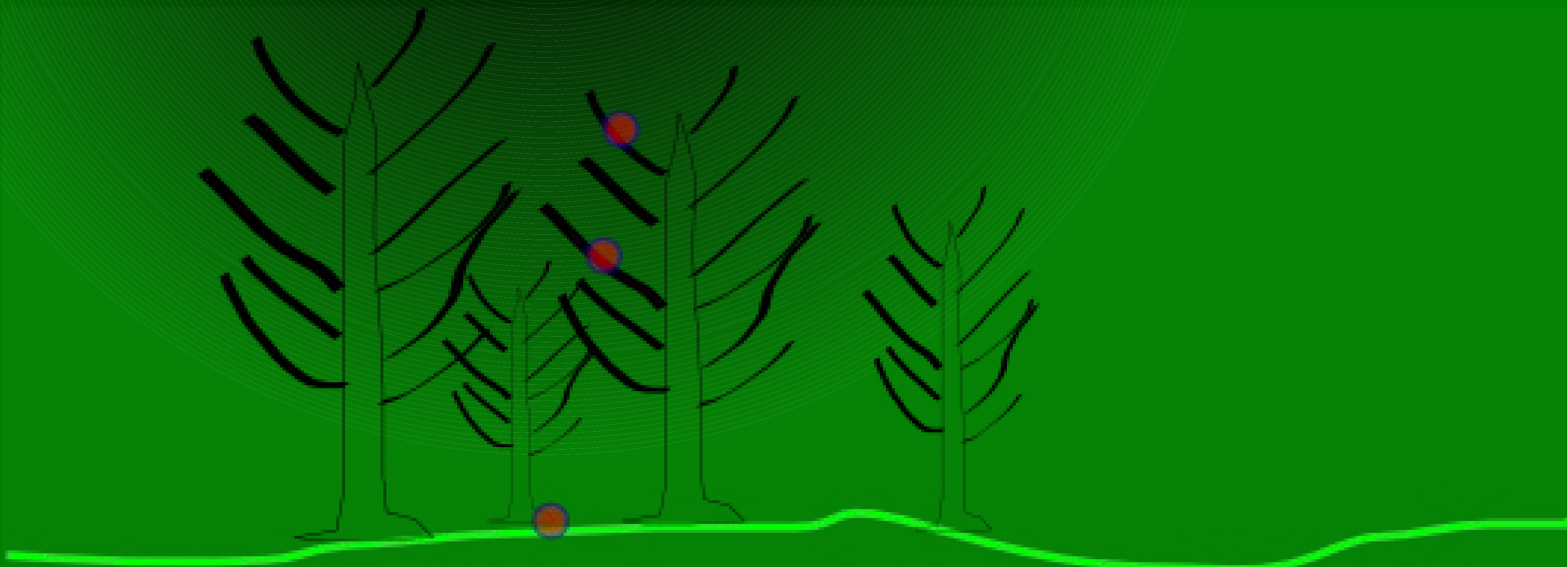
Ground Surface

LiDAR Beam Spread

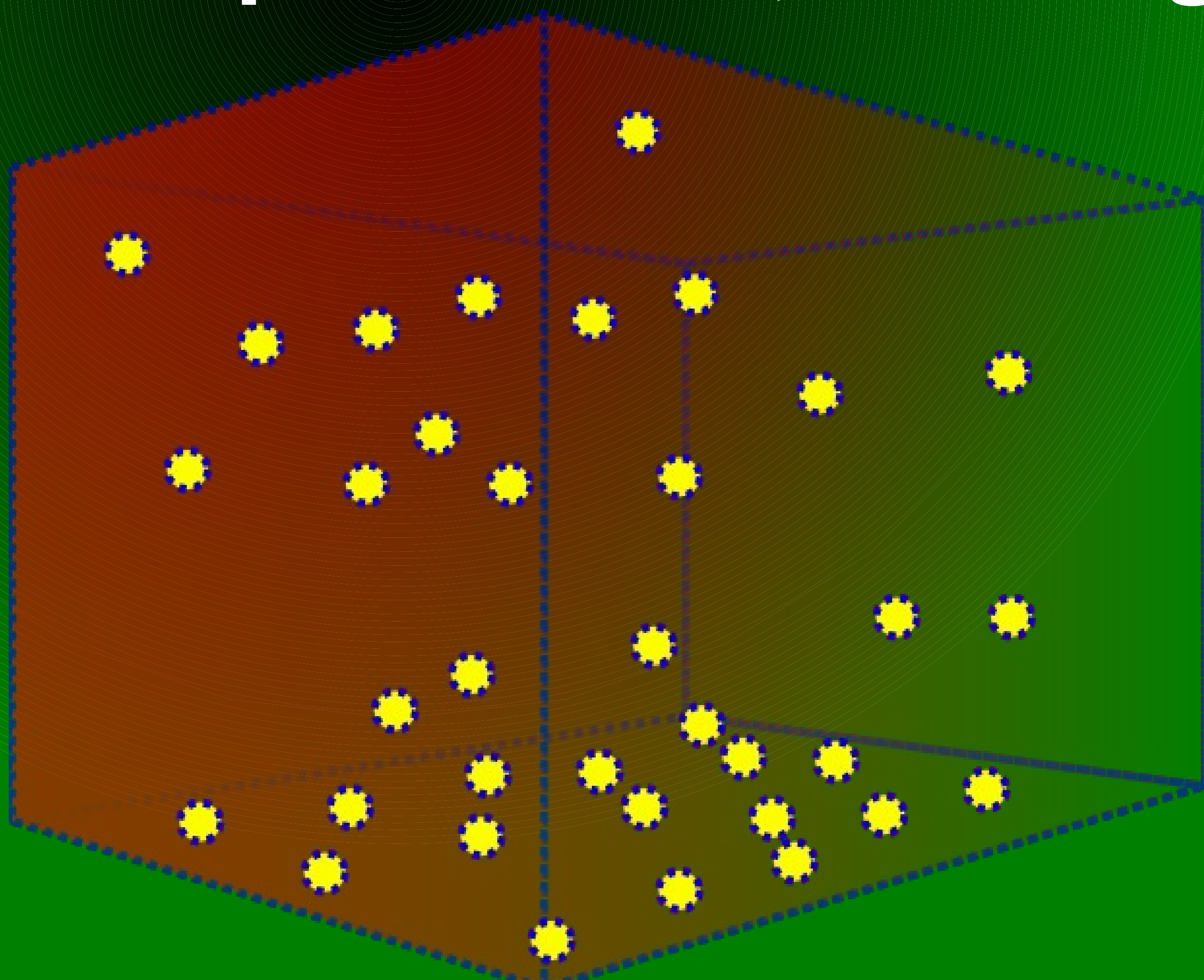
LiDAR Pulse Partial Returns

As the light beam travels down to the ground, it spreads out slightly and can encounter obstructions on the way. Power lines, rooftops, and tree branches can all give return reflections. There can be up to 5 returns per light pulse, but 1-3 is more common

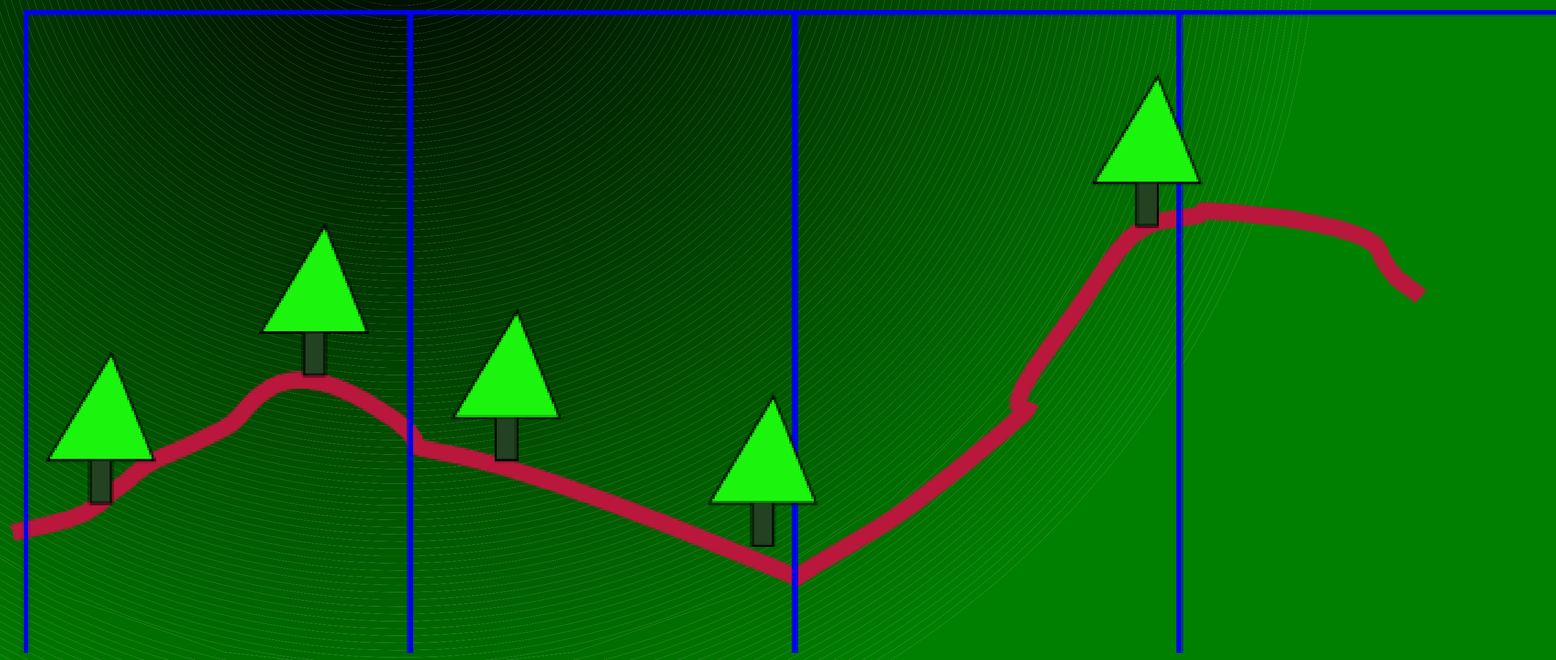
Series of X,Y, Z Points derived



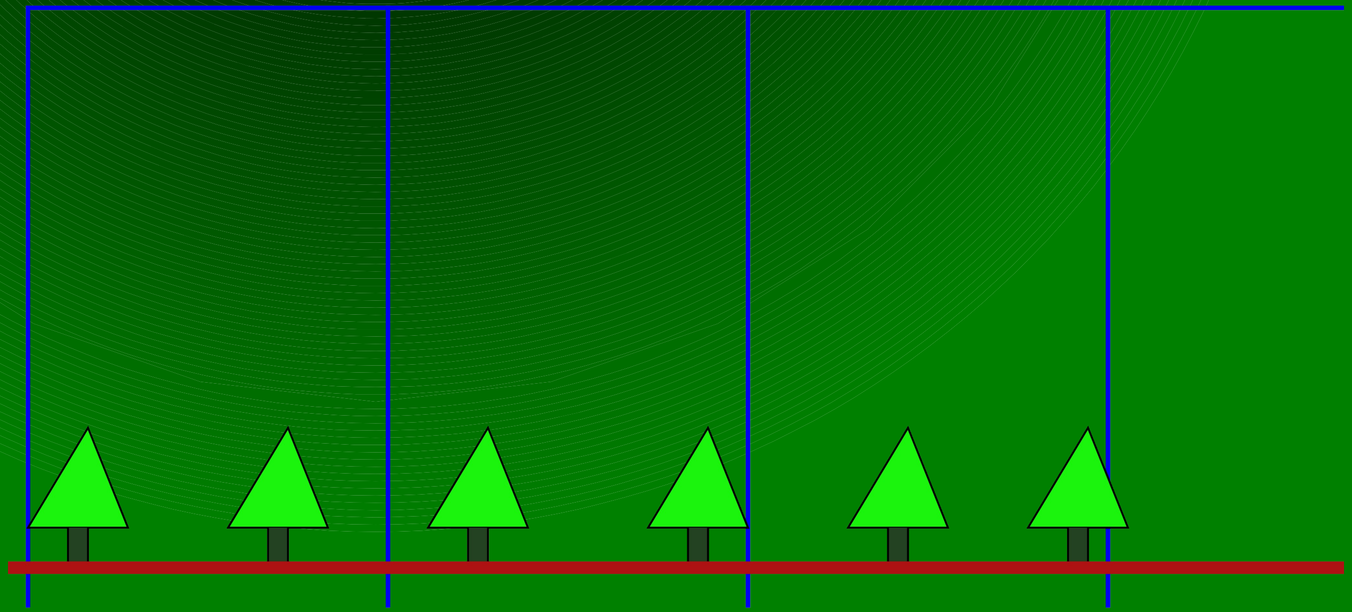
All points together give a point cloud designating reflections from objects (trees, birds, powerlines, buildings, etc)



Elevation changes within the 60 ft grid cell can make the “canopy height” artificially high!



Recomputing the Z values of the Lidar points to heights relative to the ground surface computationally “flattens” the ground for more accurate canopy height calculation and allows for different statistical analysis per grid cell



How do we get there?

- LiDAR Data was in 3 different formats , ASCII X,Y,Z and 2 binary.
- Converted all data to ASCII X, Y, Z data 4 years ago – 1 single 25.5 billion point 703 GB file. Somewhat cumbersome to work with.
- Converted to LAS format using liblas with python script into seven ~ 3.3 billion point LAS files.
- Used liblas and gdal in python to “normalize” the LAS point data so that the Z value is relative to the ground.

Python script for normalization to elevation grid

```
#!/usr/bin/python
import os,string,glob,re,gdal
from liblas import file
from liblas import header
from liblas import point
from gdalconst import *
h=header.Header()
print "/gisdata2/raster/All_1.las\n"
infile=raw_input("Enter the input lidar data points file: ")
imgfile="/gisdata2/raster/allnc_20ft_el.img"
#print "suggest /gisdata2/raster for output dir\n"
inarr=infile.split('.')
outfil=inarr[0]+"_norm.las"
#outfil=raw_input("Enter output text file name: ")
l=file.File(infile,mode='r')
lout=file.File(outfil,mode='w',header=h)
# register all of the drivers
gdal.AllRegister()
ds=gdal.Open(imgfile,GA_ReadOnly)
if ds is None:
    print 'Could not open image'
    sys.exit(1)
# get image size
rows = ds.RasterYSize
cols = ds.RasterXSize
bands = ds.RasterCount
# get georeference info
transform = ds.GetGeoTransform()
xOrigin = transform[0]
yOrigin = transformAsArray(xOffset, yOffset, 1, 1)
pixelWidth = transform[1]
pixelHeight = transform[5]
for p in l:
    x=float(p.x)
    y=float(p.y)
    z=float(p.z)
    # compute pixel offset
    xOffset = int((x - xOrigin) / pixelWidth)
    yOffset = int((y - yOrigin) / pixelHeight)
    band = ds.GetRasterBand(1) # 1-based index 0? 1?
    data = band.Readr(value) :continue
    value = data[0,0]
    #print value,"11","\n"
    if "nan" in st[3]
    znorm = z-value
    #print znorm,"\n"
    pt=point.Point()
    pt.x=p.x
    pt.y=p.y
    pt.z=znorm
    lout.write(pt)

l.close()
lout.close()
#25561312019 points in allreturns
```

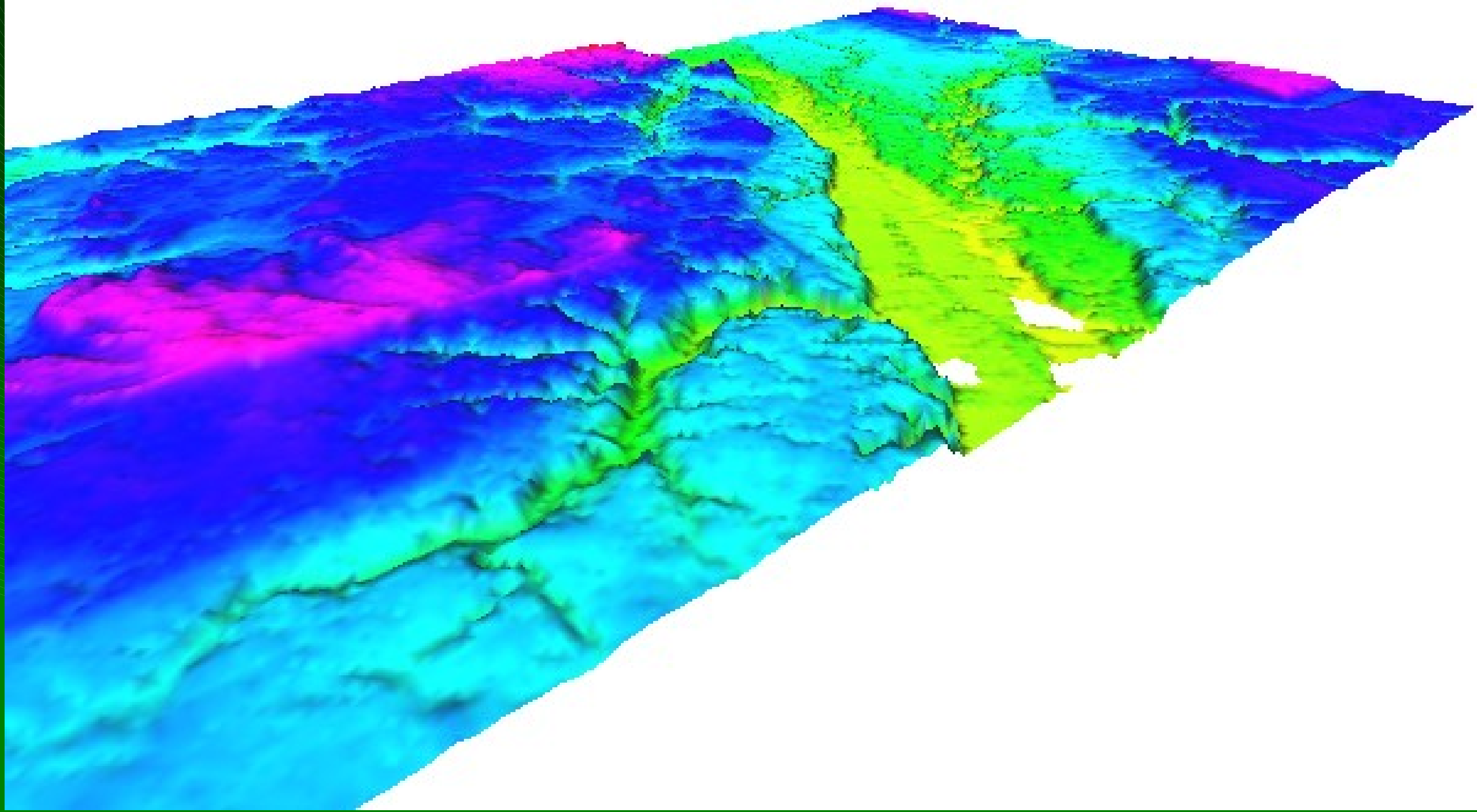
Processing in Bulk

- Processing performed on a Dual Quad core 2Ghz Xeon server with 42 GB RAM, running 64 bit Ubuntu 11.10 Linux using GRASS70 (compiled with liblas library)
- Used r.in.lidar , http://grass.osgeo.org/grass70/manuals/r.in.lidar.html, to perform basic per grid cell statistics on Z values of points.
- Analysis could be performed simultaneously on 7 cores of the computer (one for each core) with per process memory demand ranging from 4.5 GB to 20 GB of RAM per process.

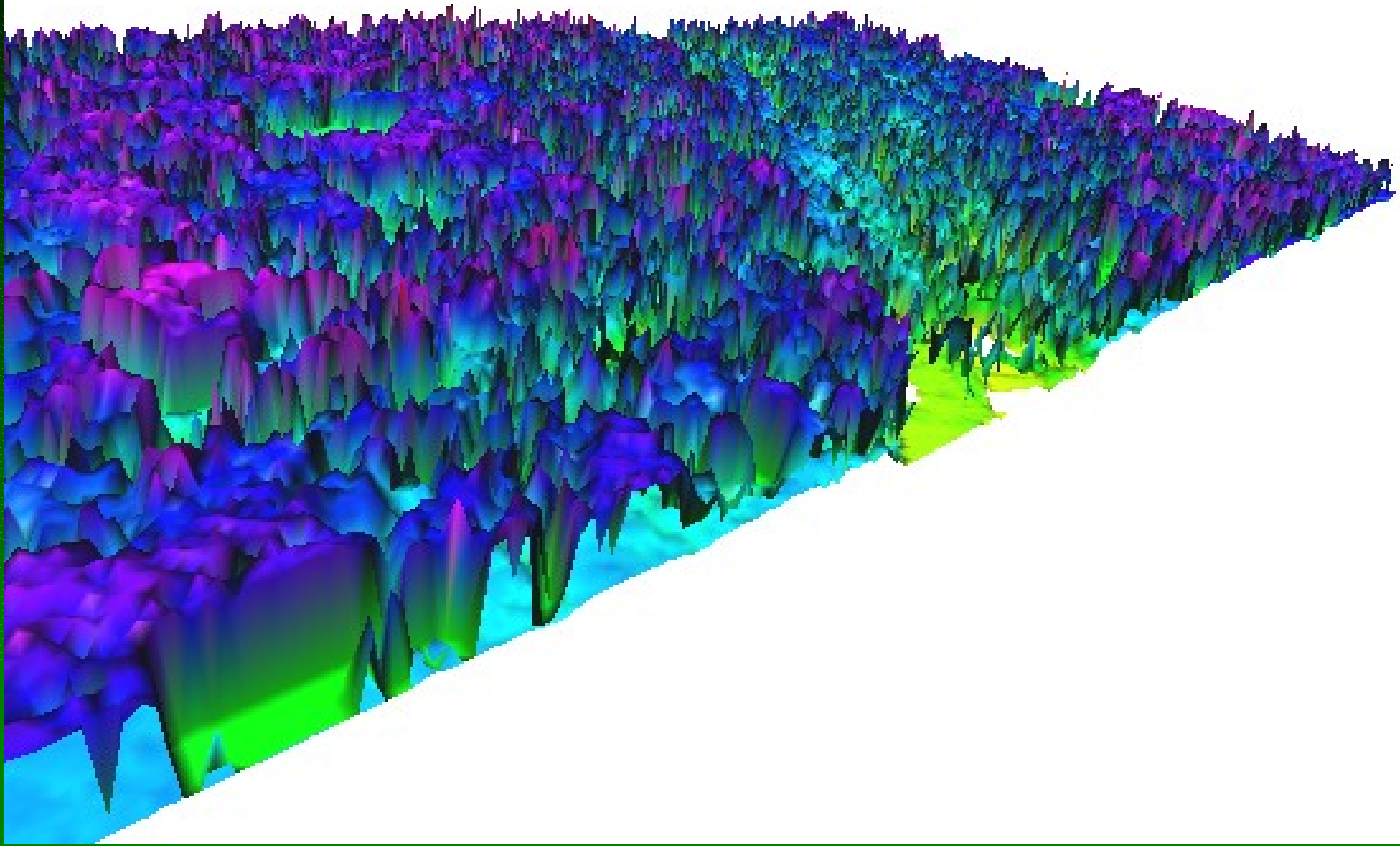
Processing in Bulk - continued

- Analysis performed include range,skewness, n,max, variance, and coefficient of variance, and standard deviation.
- Z values below -10ft below ground and above 250ft above ground were excluded from calculation.
- The memory demand of skewness analysis (20 GB of RAM with 30% of the map in memory at a time) required that only 2 skewness analysis be run simultaneously

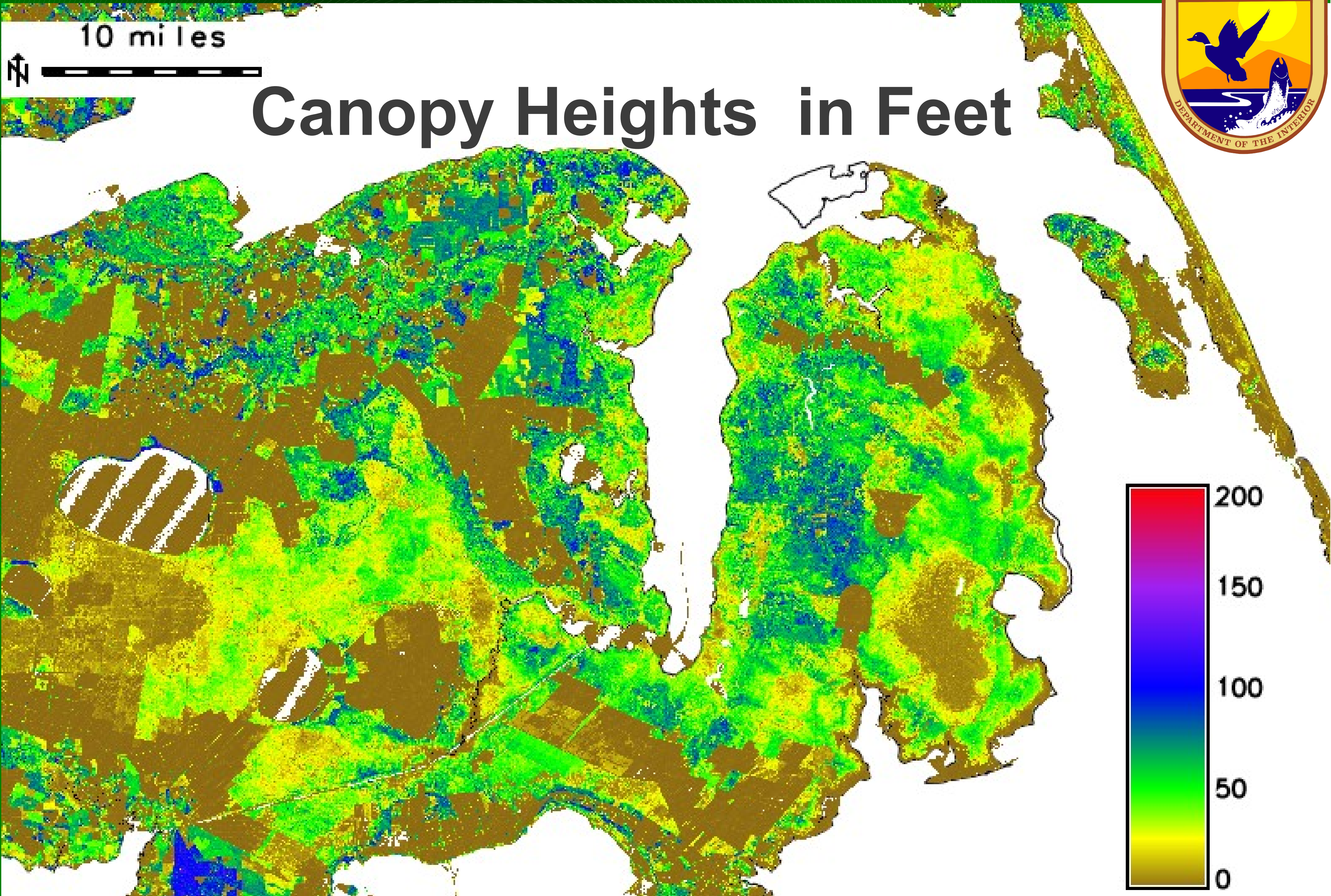
Derive Land Surface from Ground Points



Difference Between Ground Surface and First Returns is Canopy Height

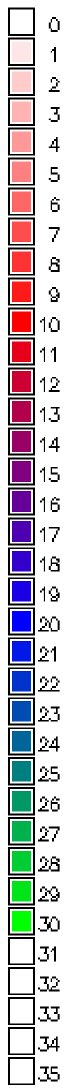


Canopy Heights in Feet



Map Display 1

Cross Flights and overlaps can make raw cell counts useless, so statistical measures that allow for comparison between cells with different point densities are better for structure



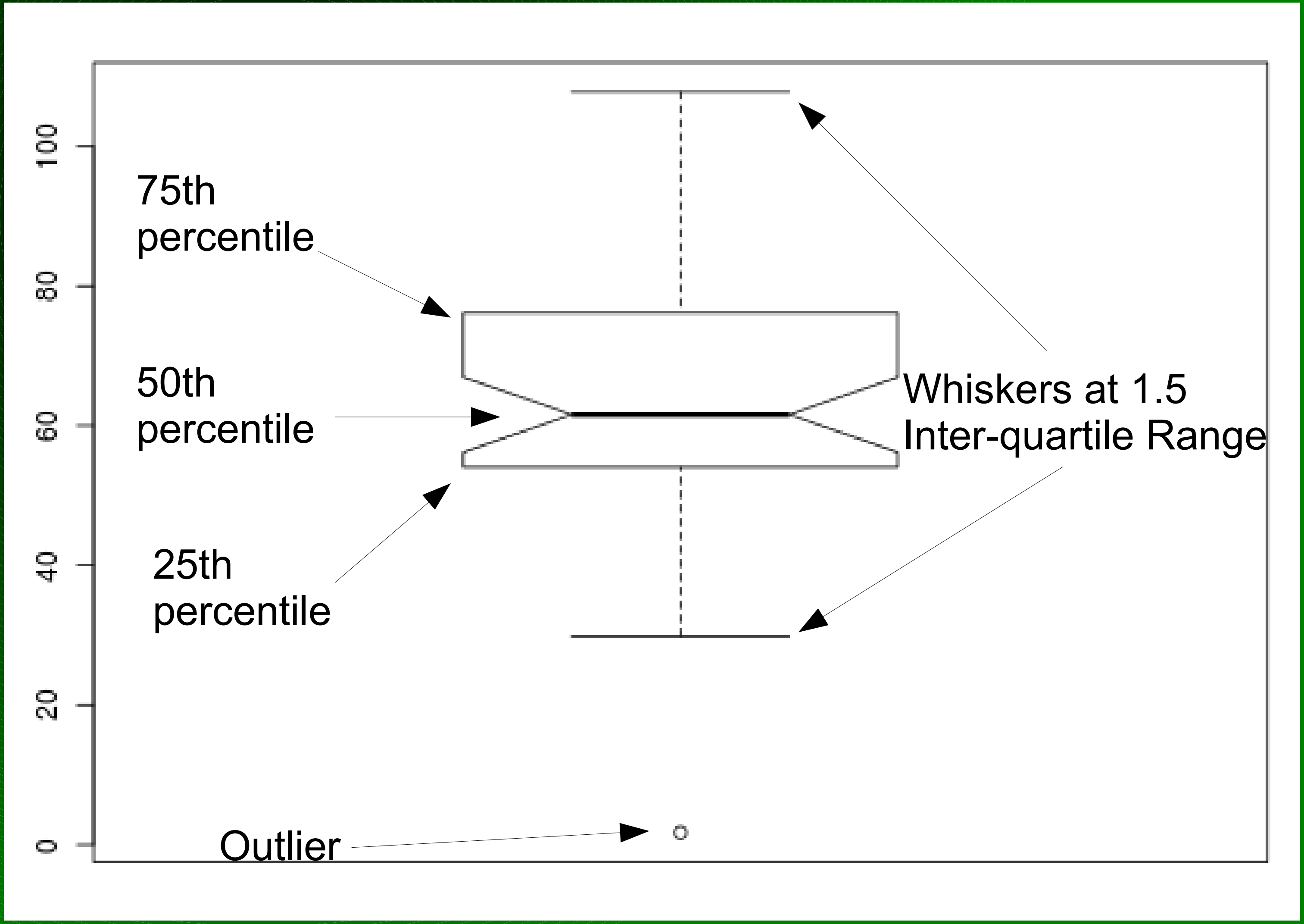
Skewness , Max Height, Mean Height,and Percentages of Points by Layer:

Buffer Point Locations of Bird Species by 25m,50m, and 75m. (N for RCW =702, rest of species < 20)

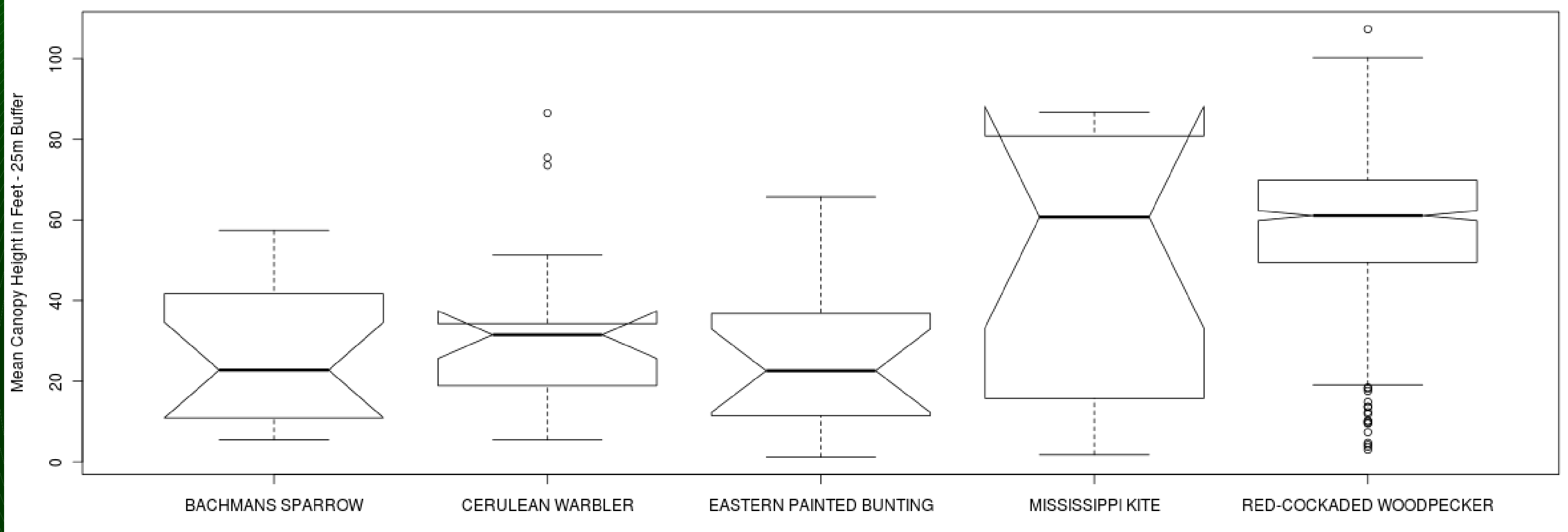
Create Zonal statistics for each buffered polygon for each raster layer.

Throw into R to see what patterns show up.

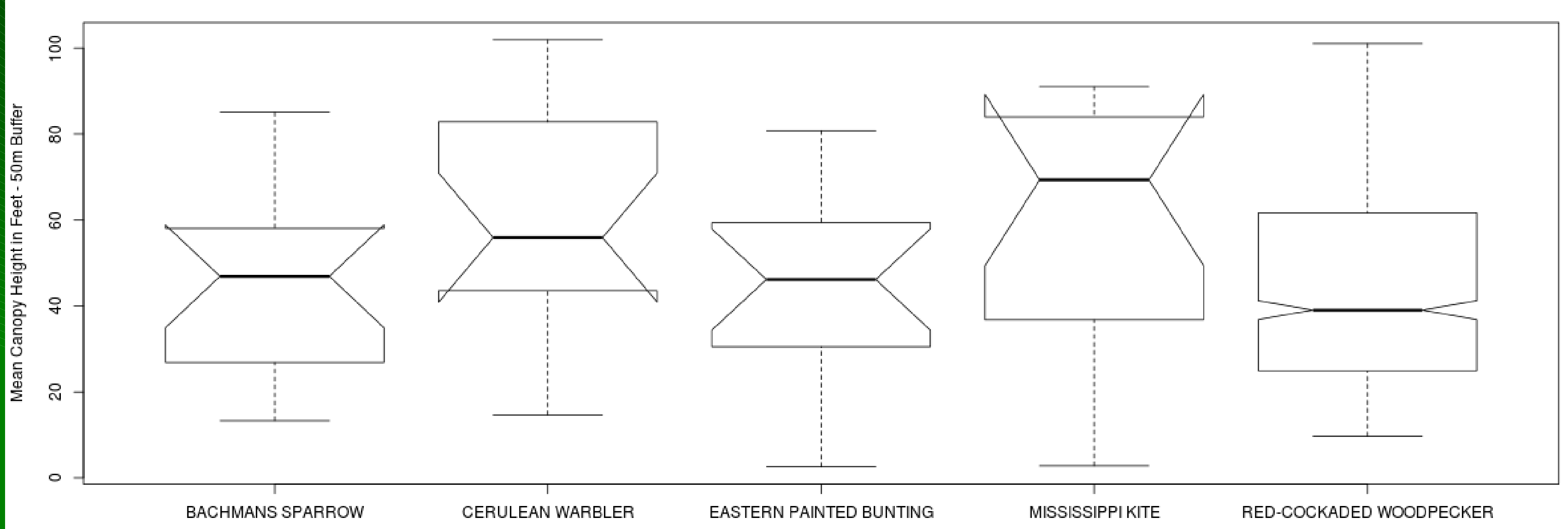
Notched Box plots in R

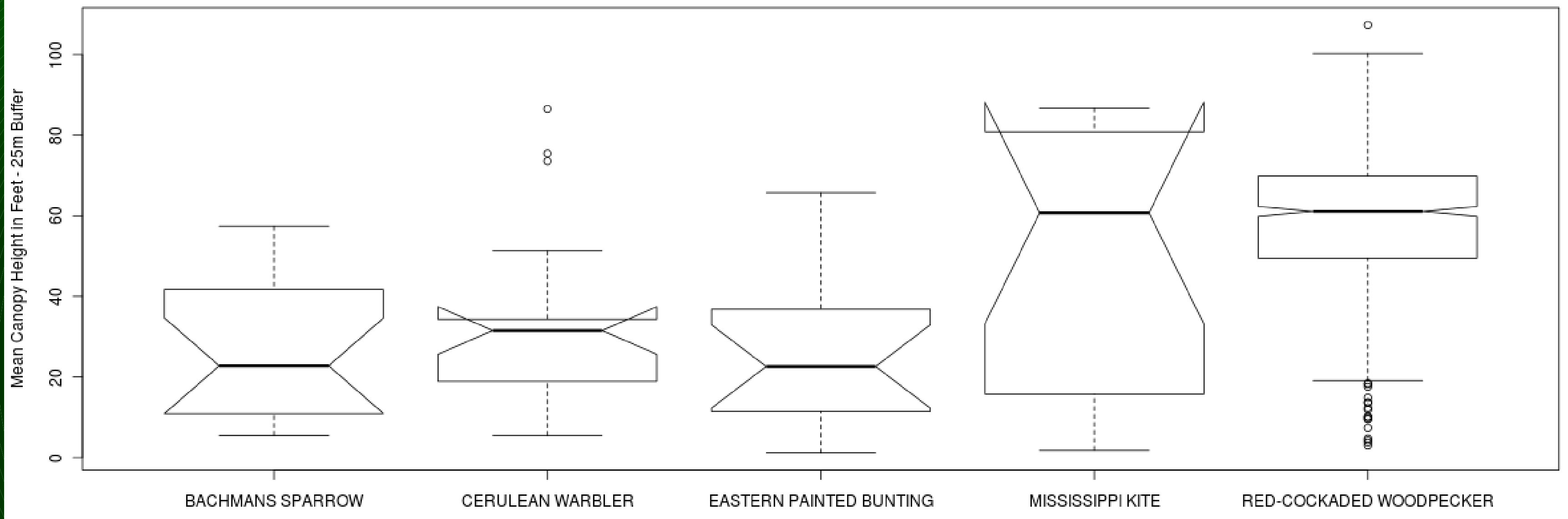


The width of the notches is proportional to the inter-quartile range of the sample and inversely proportional to the square root of the size of the sample. The whiskers extend about 1.5 times the length of the box away from the box. Data outside of that distance are represented separately as outlying points.

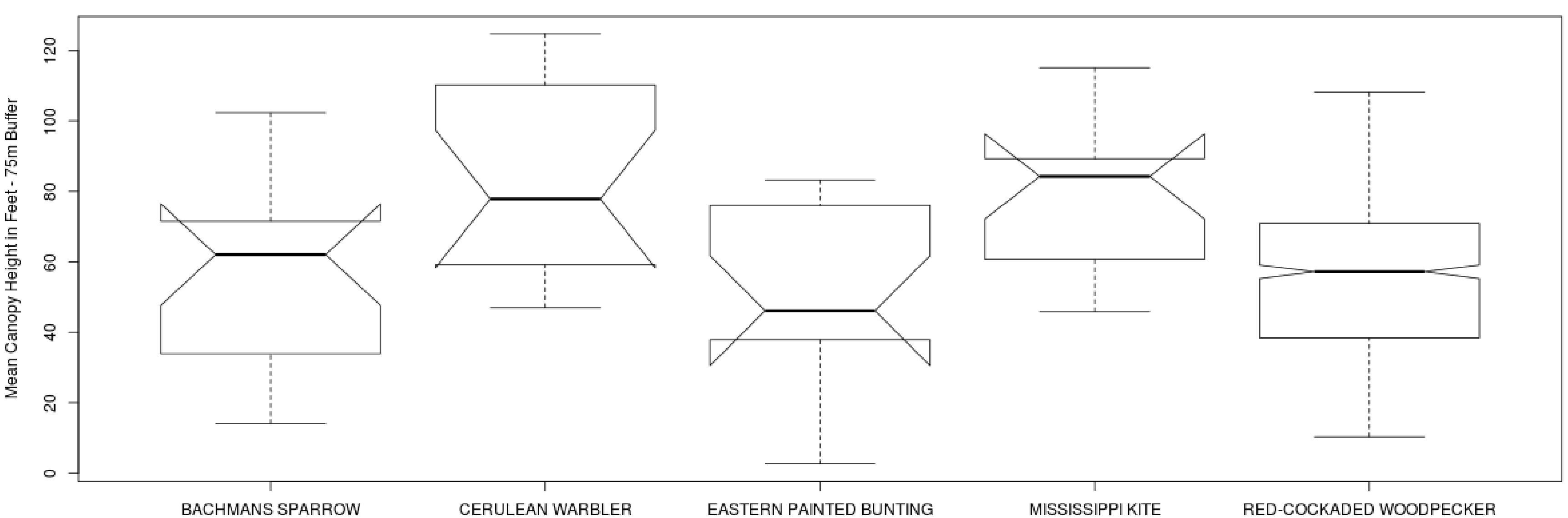


Mean Height seems to be a useful measure at 25m buffer with differences at 50m buffer . Many C. Warbler sites riverside.

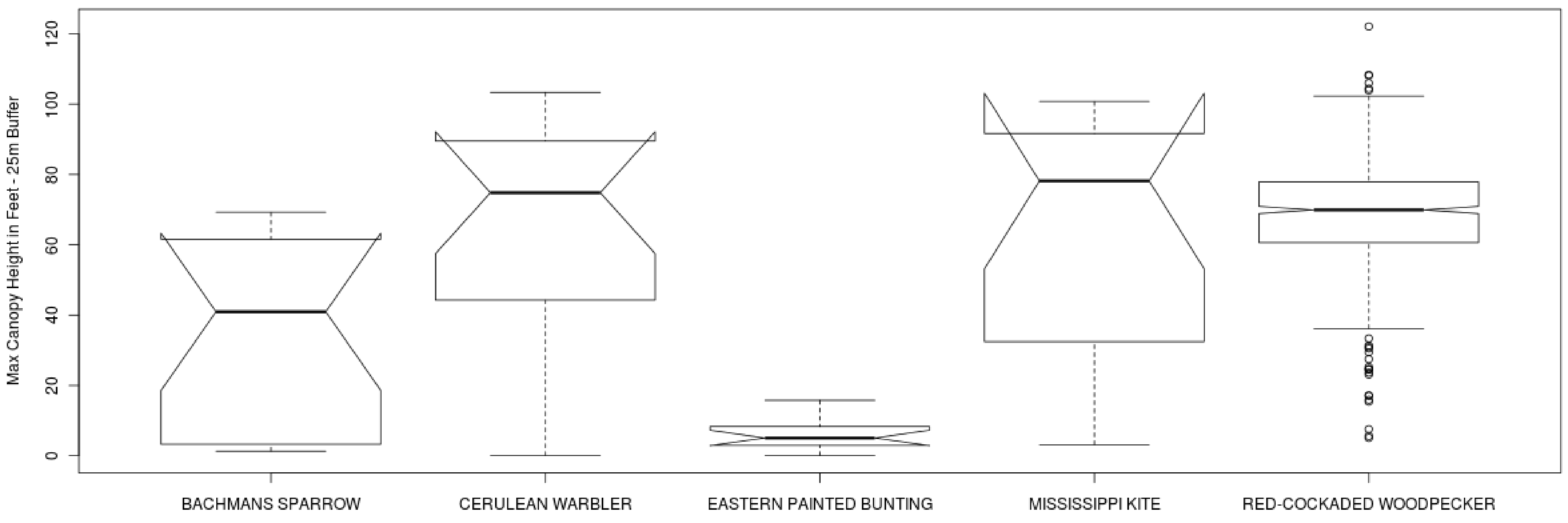




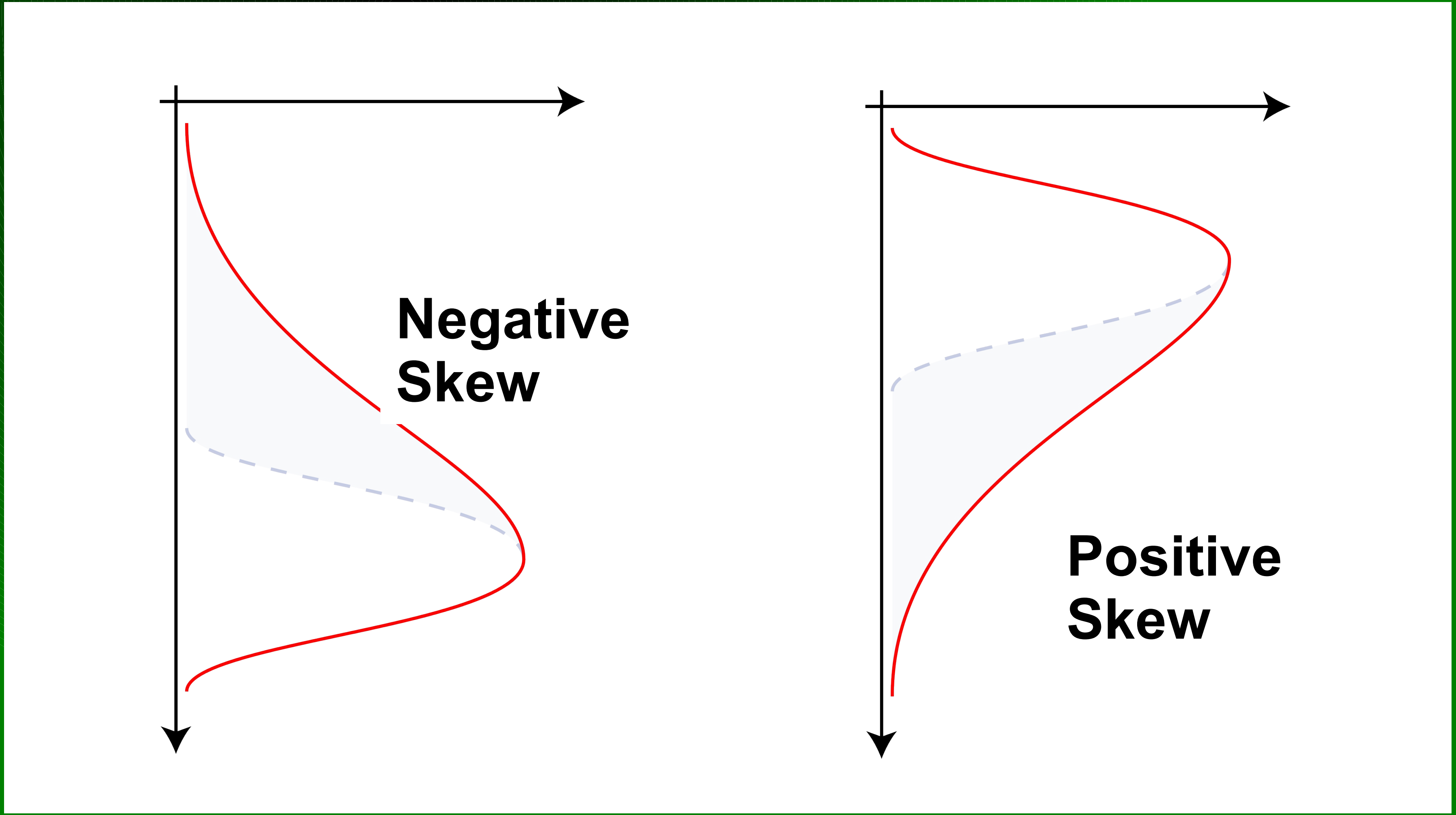
Mean Height at 75m seems to be washing out differences between species, see RCW and Painted Bunting



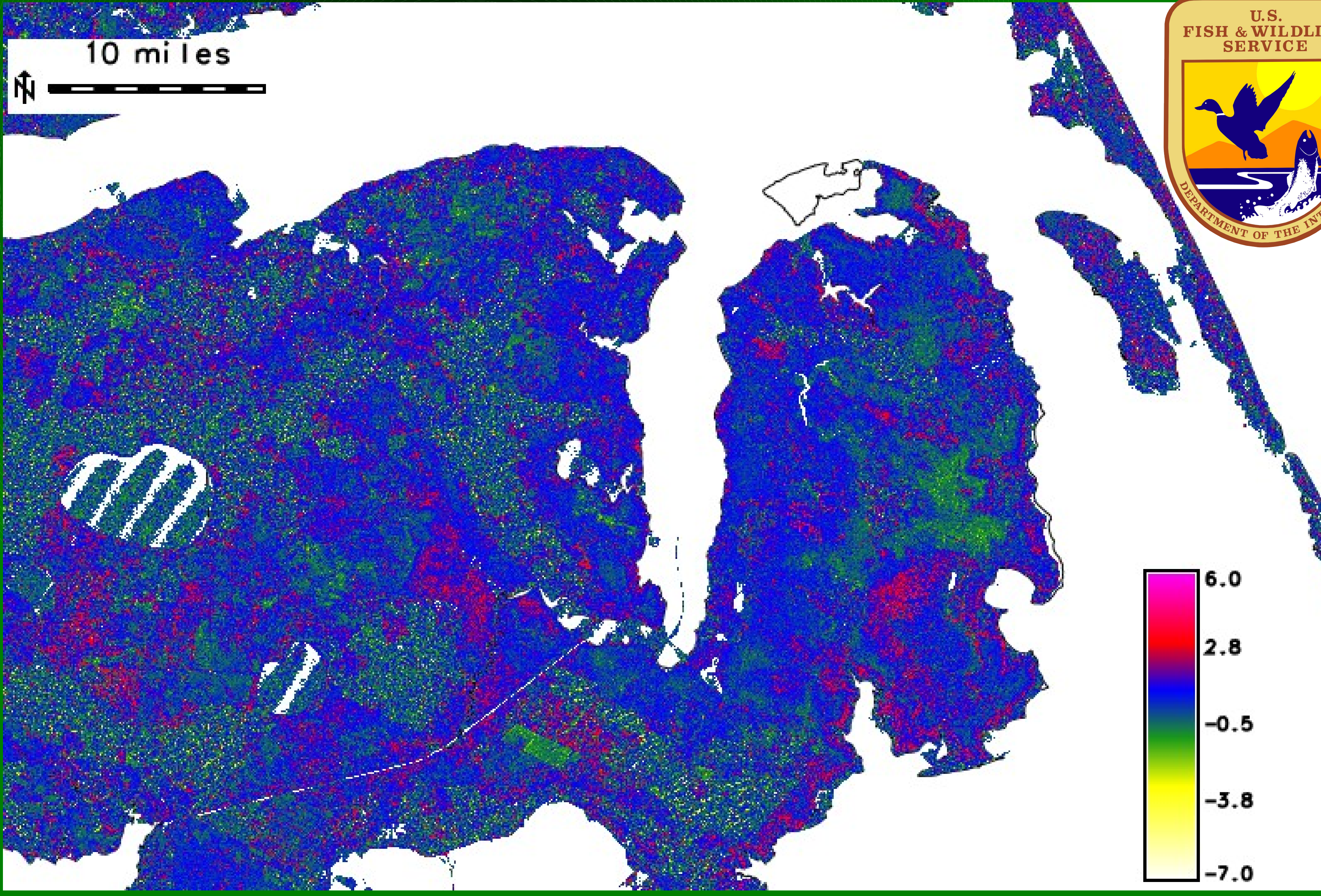
Max Height at 25m



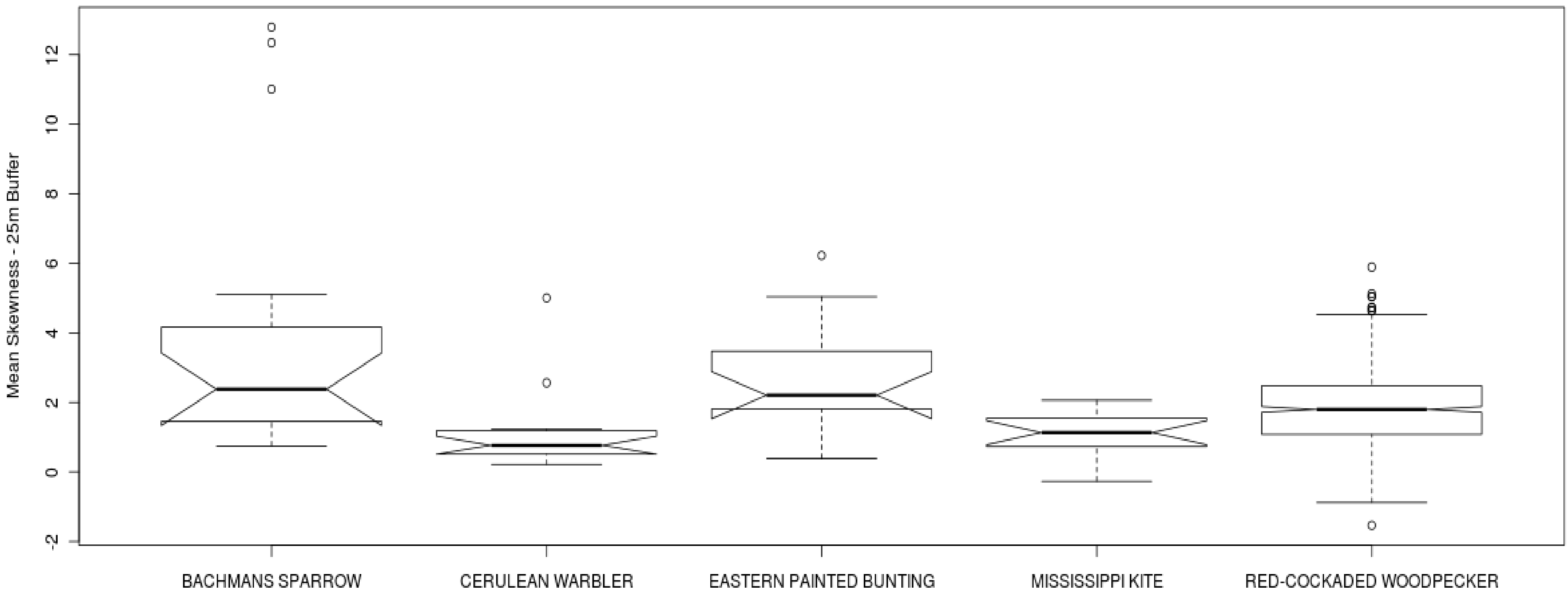
Skewness of Z values of LiDAR points in each cell.



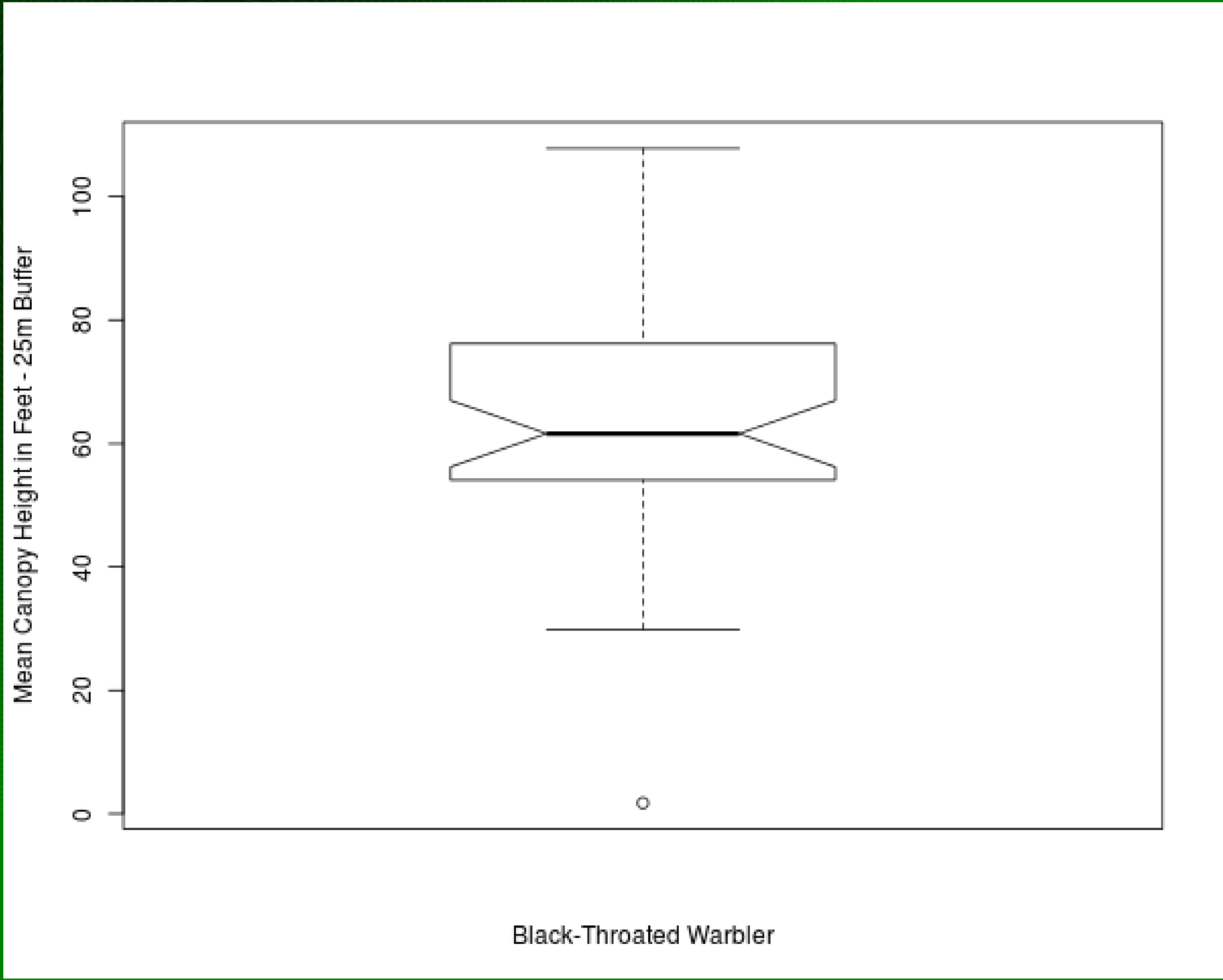
Skewness of Z values of LiDAR points .



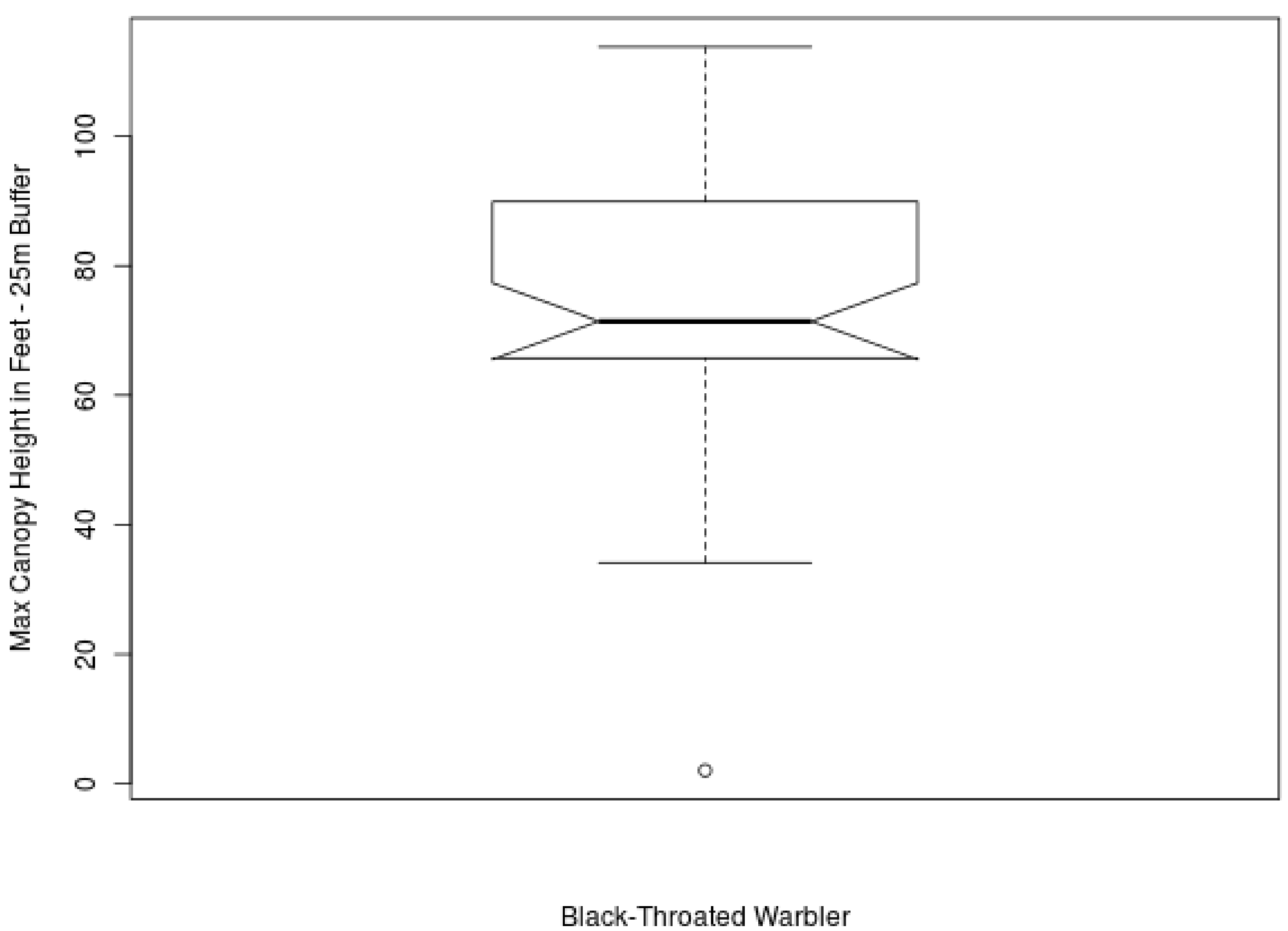
Skewness at 25m



N=43 More certainty – Mean Height

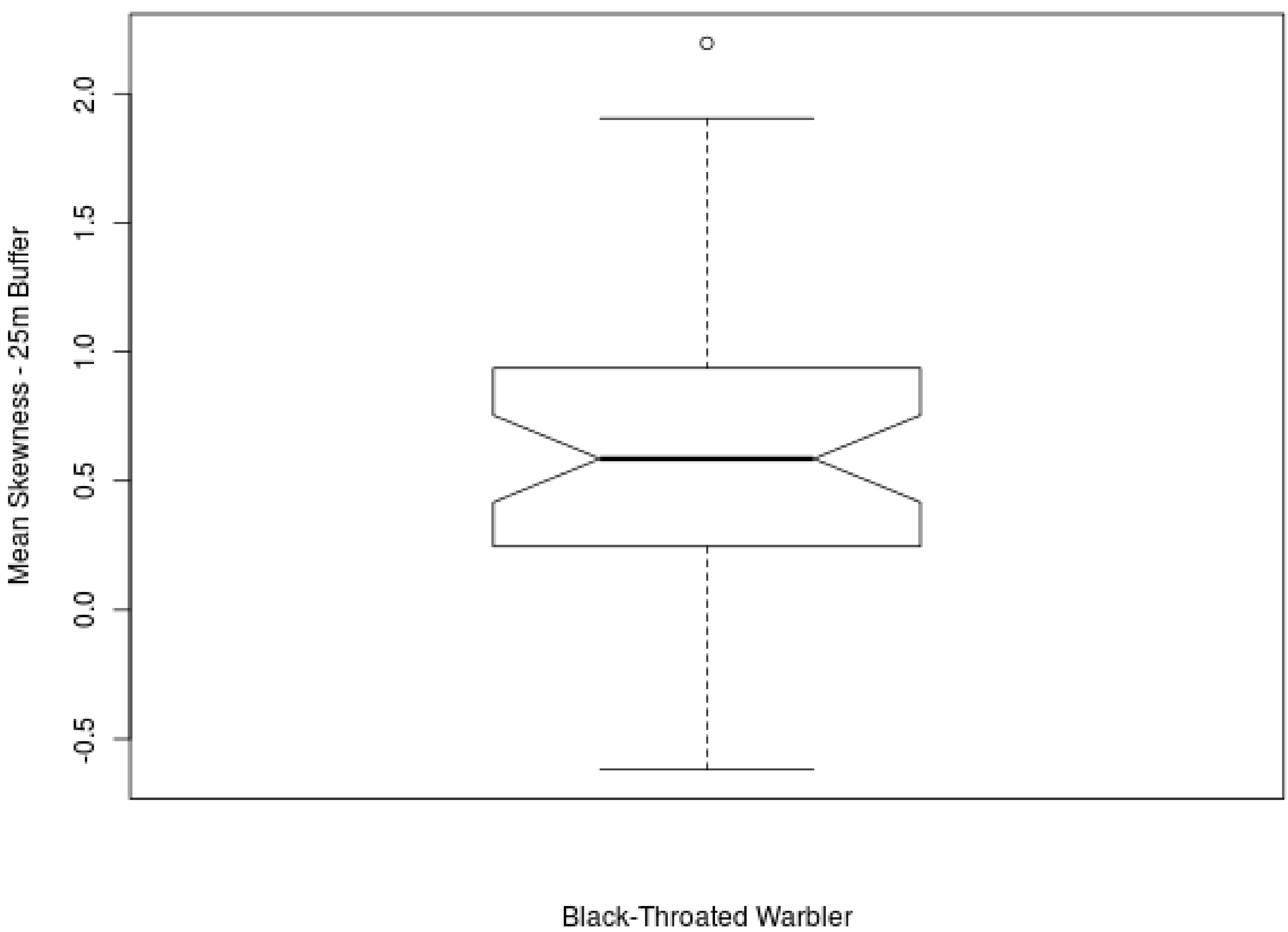


Max Canopy



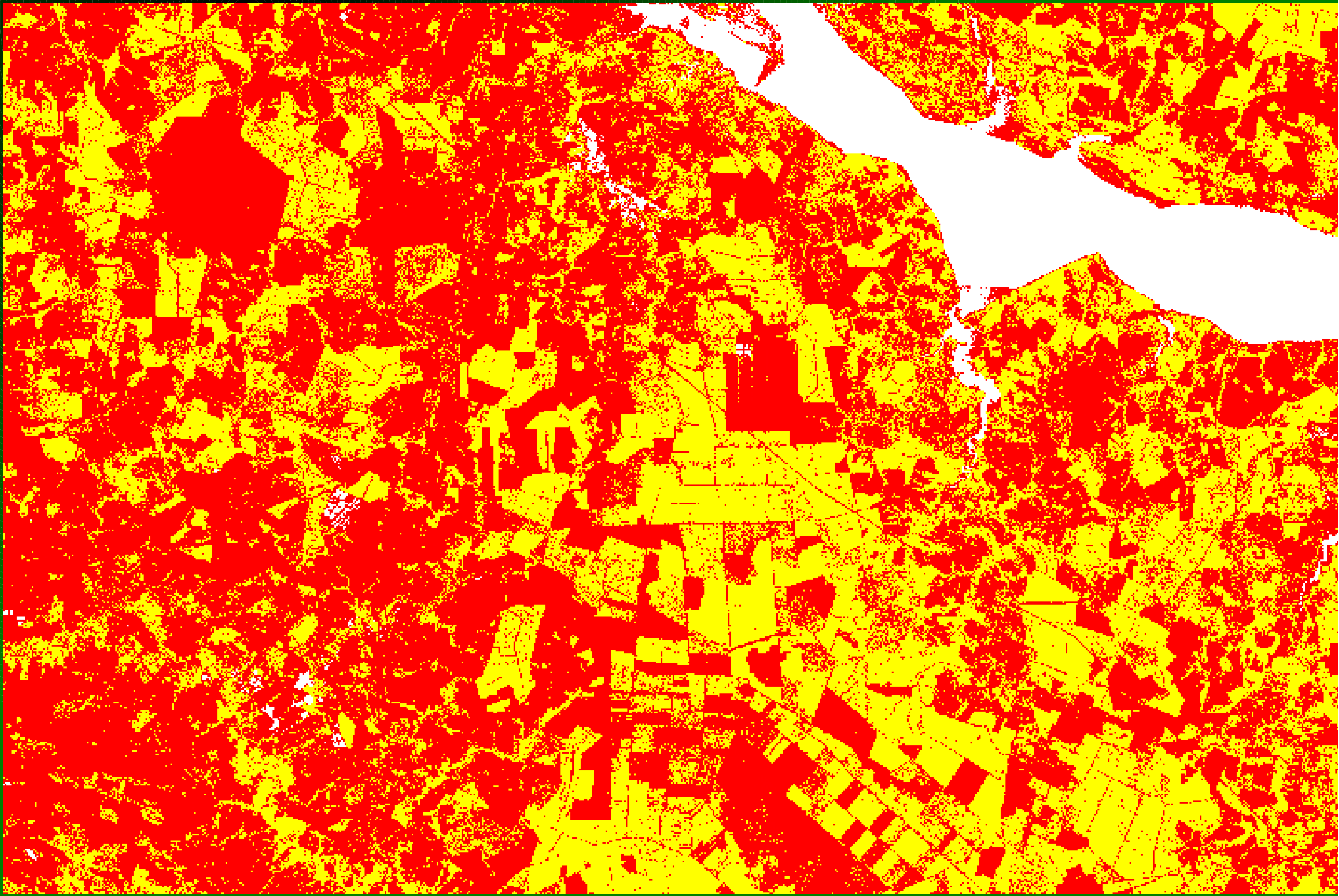
Max Canopy Height in Feet - 25m Buffer
0
20
40
60
80
100
Black-Throated Warbler

Skewness for Black – Throated Warbler



Identify probable areas of RCW occurance?

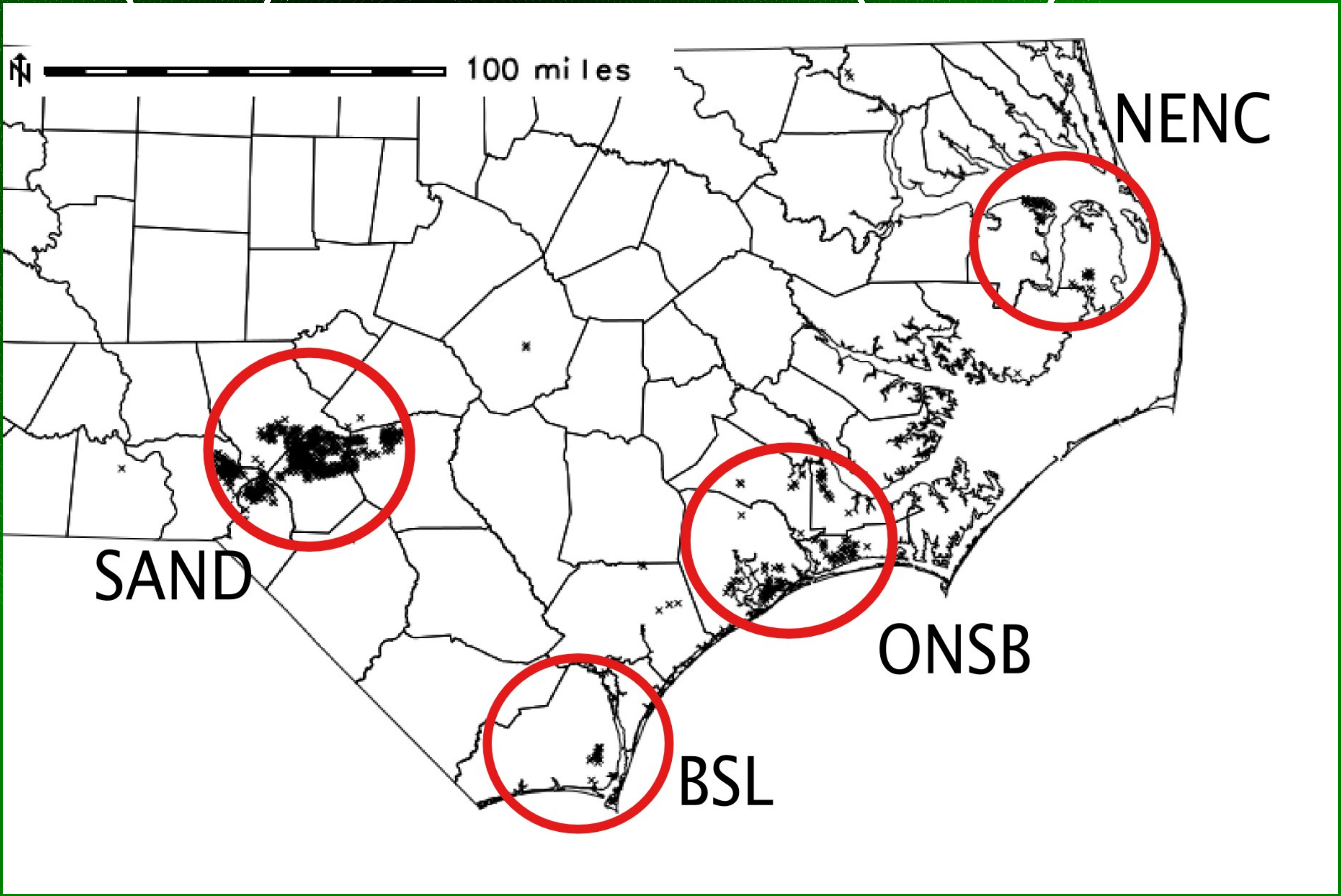
canopy height within 1 SD of the mean height for RCW



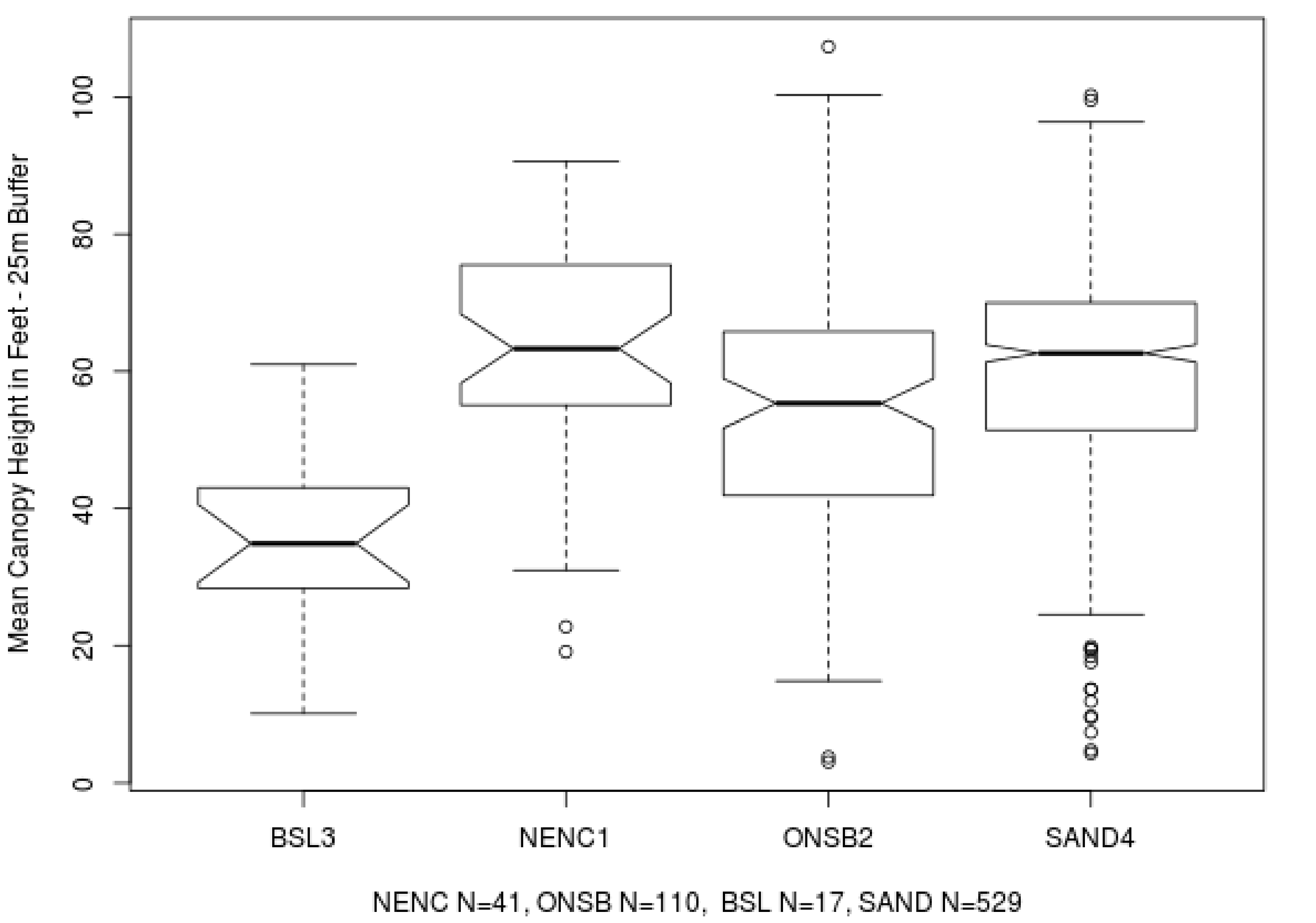
RCW canopy heights filtered by RCW skewness mask (1 SD around mean skewness for RCW)

Red-cockaded Woodpeckers (RCW) are interesting case. RCWs occur in Frequent Fire Longleaf Pine Savannahs in Sandhills of North Carolina, in smaller, slower growing Longleaf Pine in Southeastern NC and also in Pocosin swamp areas of mixed deciduous/pond pine in NE North Carolina. Are there measurable differences between the vegetation structures as well as the vegetation compostion?

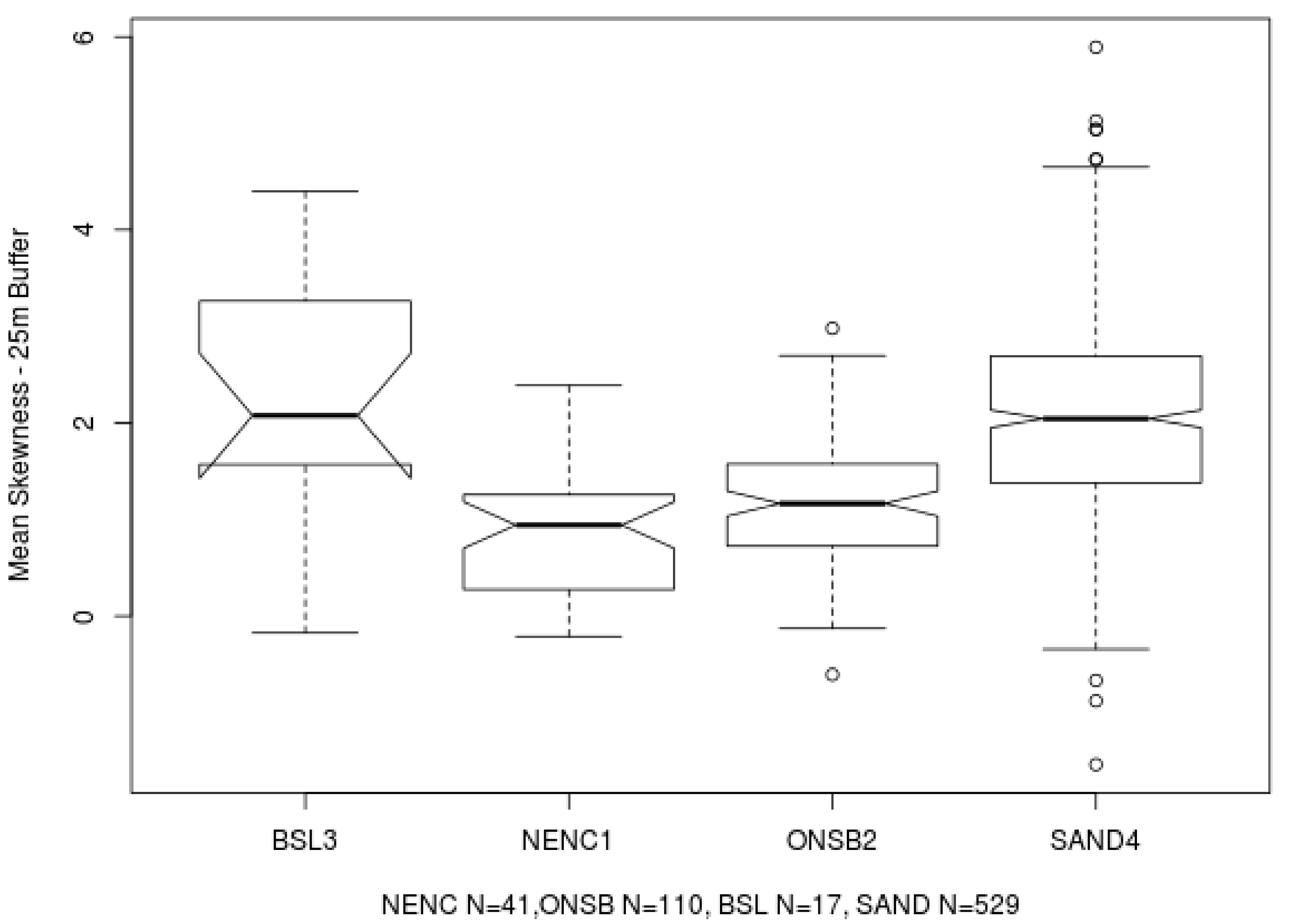
RCW Populations (1998-2003) in Northeast NC (NENC), Onslow Bight (ONSB), SE NC (BSL), and NC Sandhills (SAND)



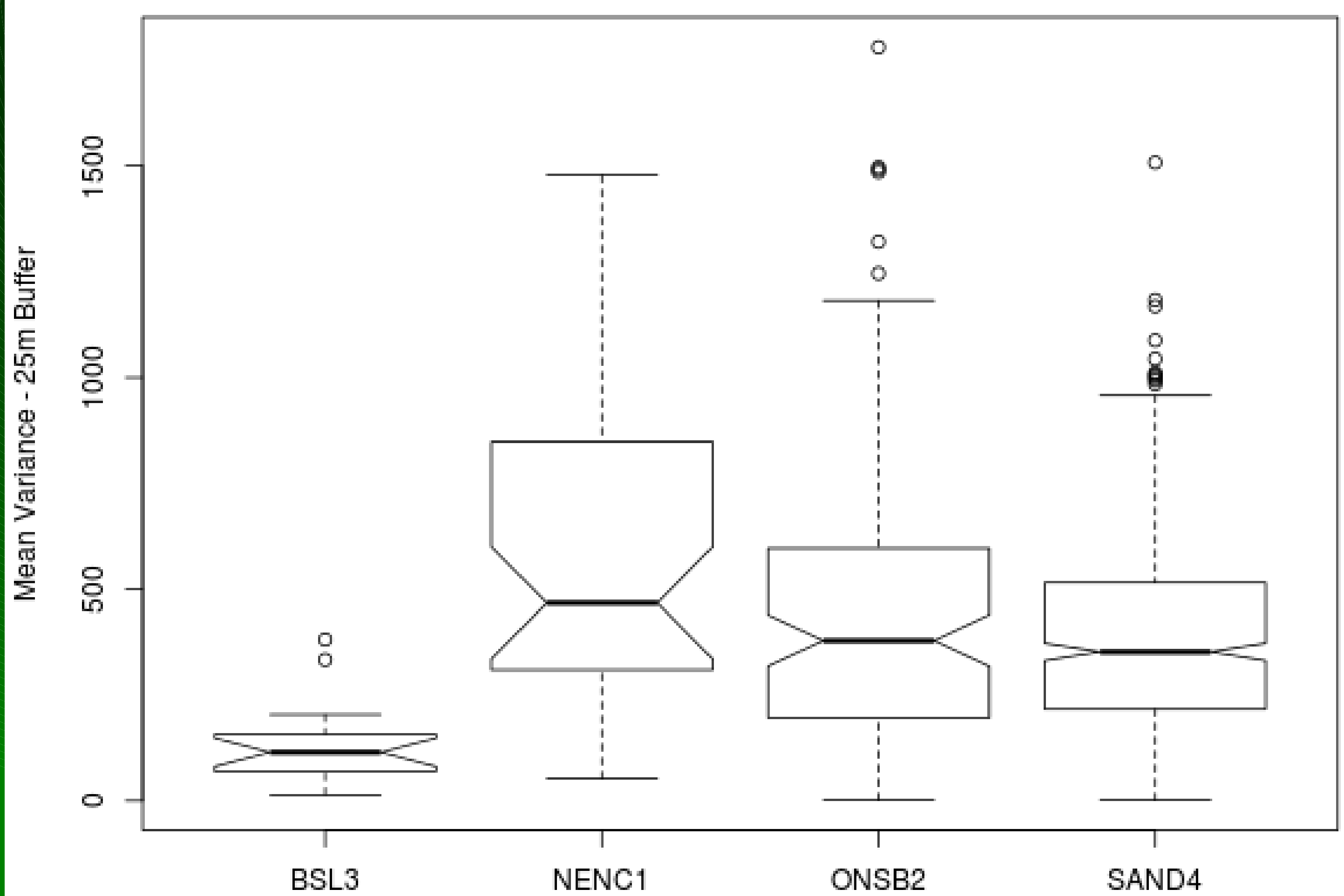
RCW Canopy Heights in Northeast NC (NENC1), Onslow Bight (ONSB2), SE NC (BSL3), and NC Sandhills (SAND4)



RCW Skew in Northeast NC (NENC1), Onslow Bight (ONSB2), SE NC (BSL3), and NC Sandhills (SAND4)

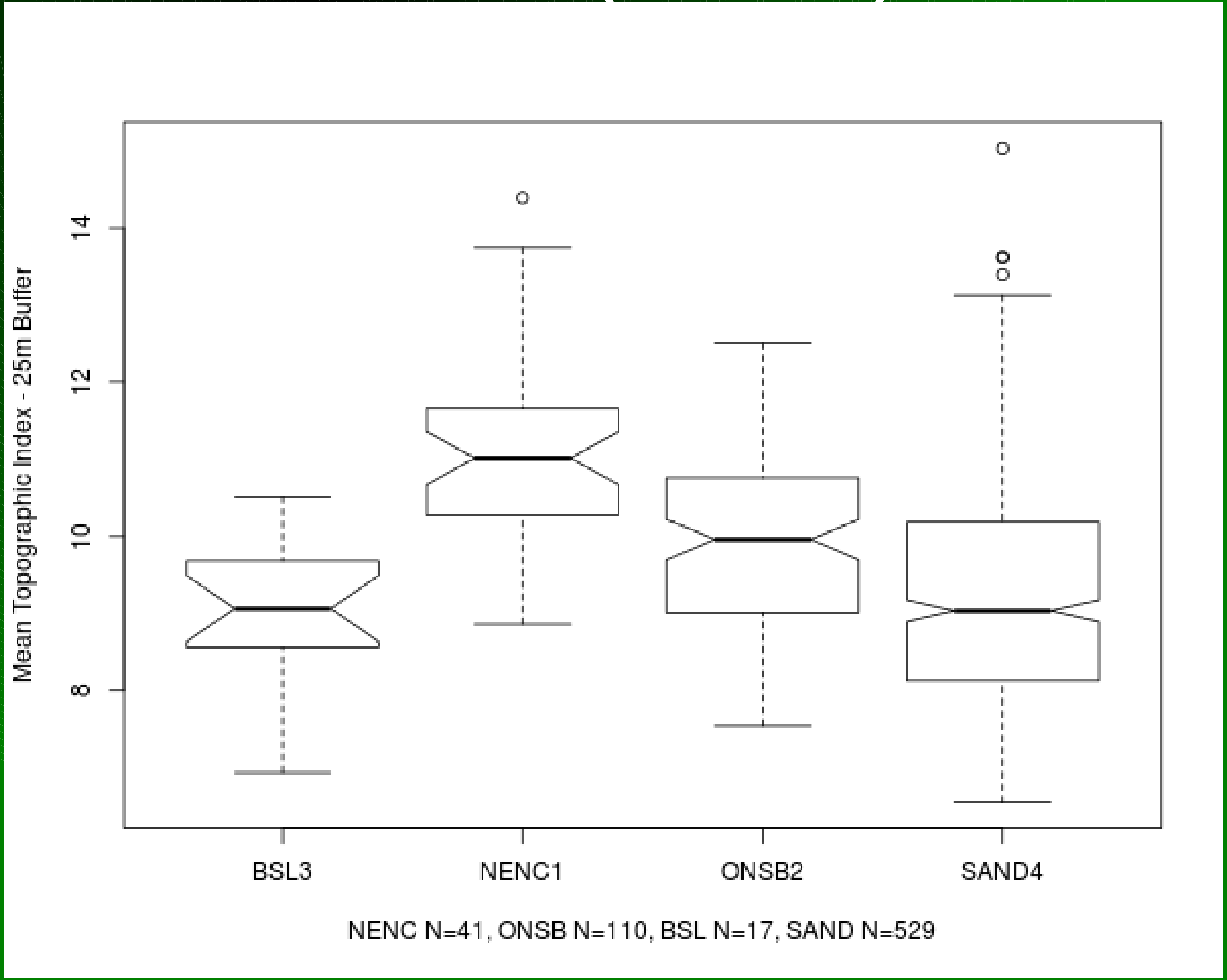


RCW Variance in Northeast NC (NENC1), Onslow Bight (ONSB2), SE NC (BSL3), and NC Sandhills (SAND4)

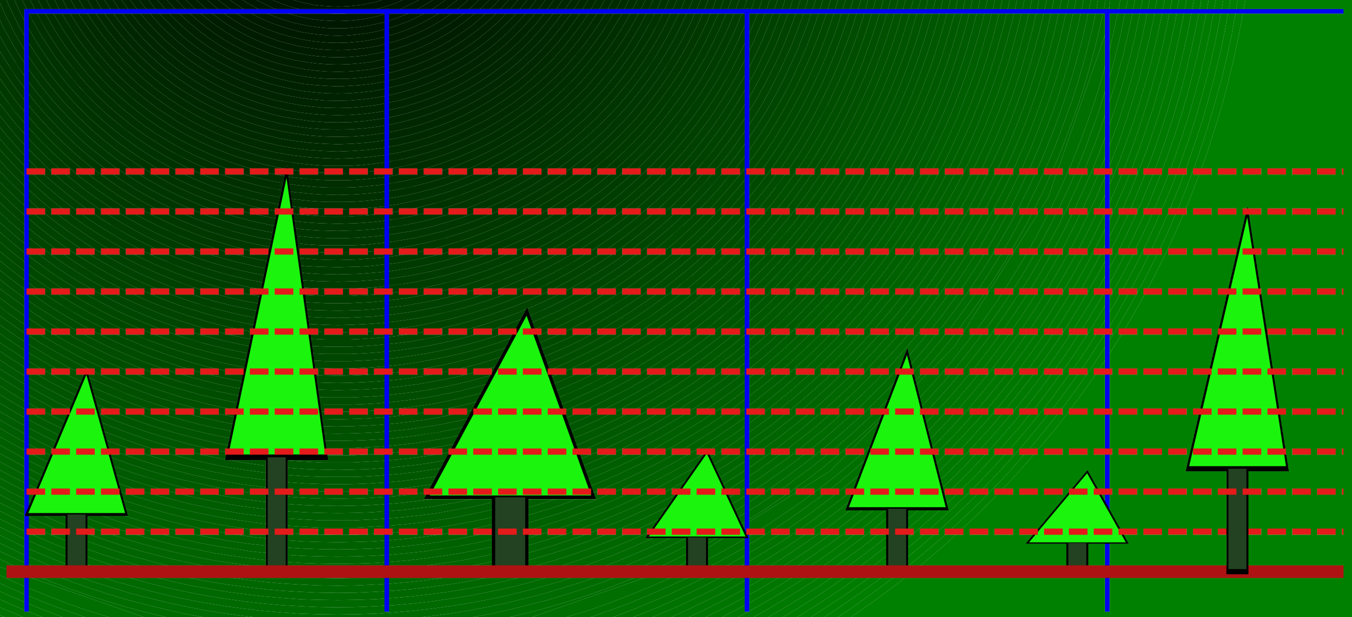


NENC N=41,ONSB N=110, BSL N=17, SAND N=529

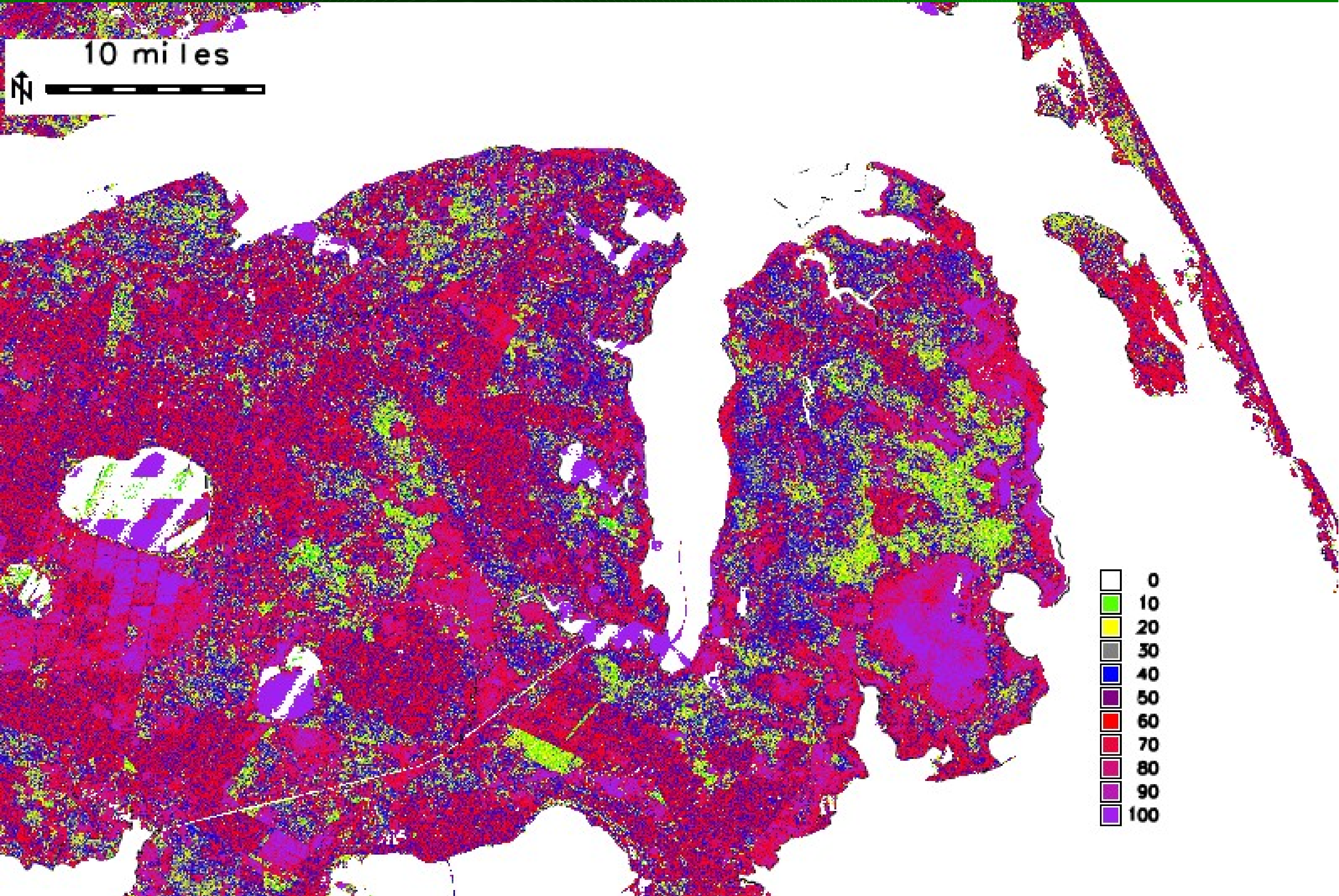
RCW Topographic Index (wetness) of ground surface in Northeast NC (NENC1), Onslow Bight (ONSB2), SE NC (BSL3), and NC Sandhills (SAND4)



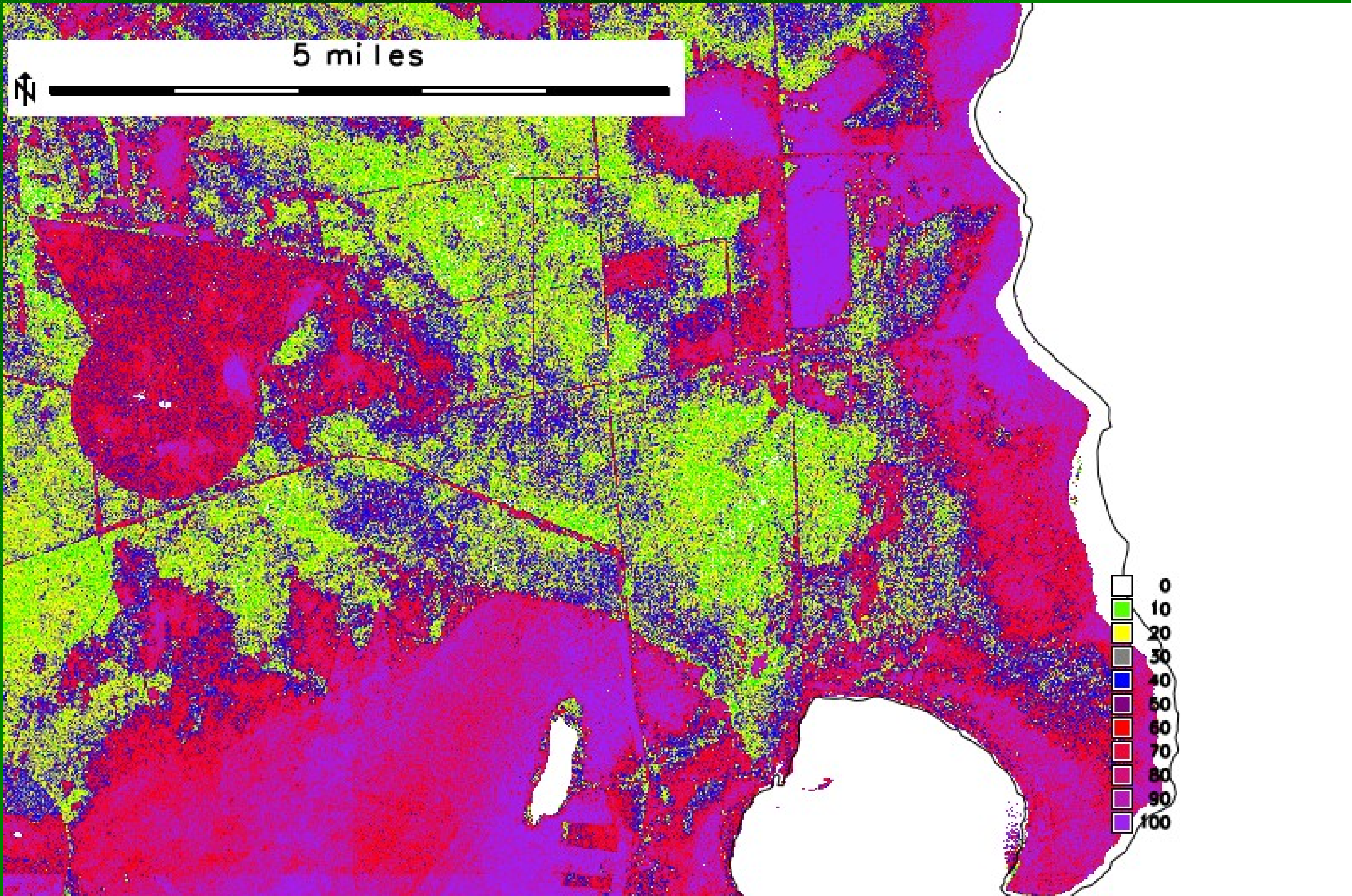
Slice the Point Cloud in Horizontal10 ft Layers and calculate the percentage of points in each grid cell that fall in each layer.



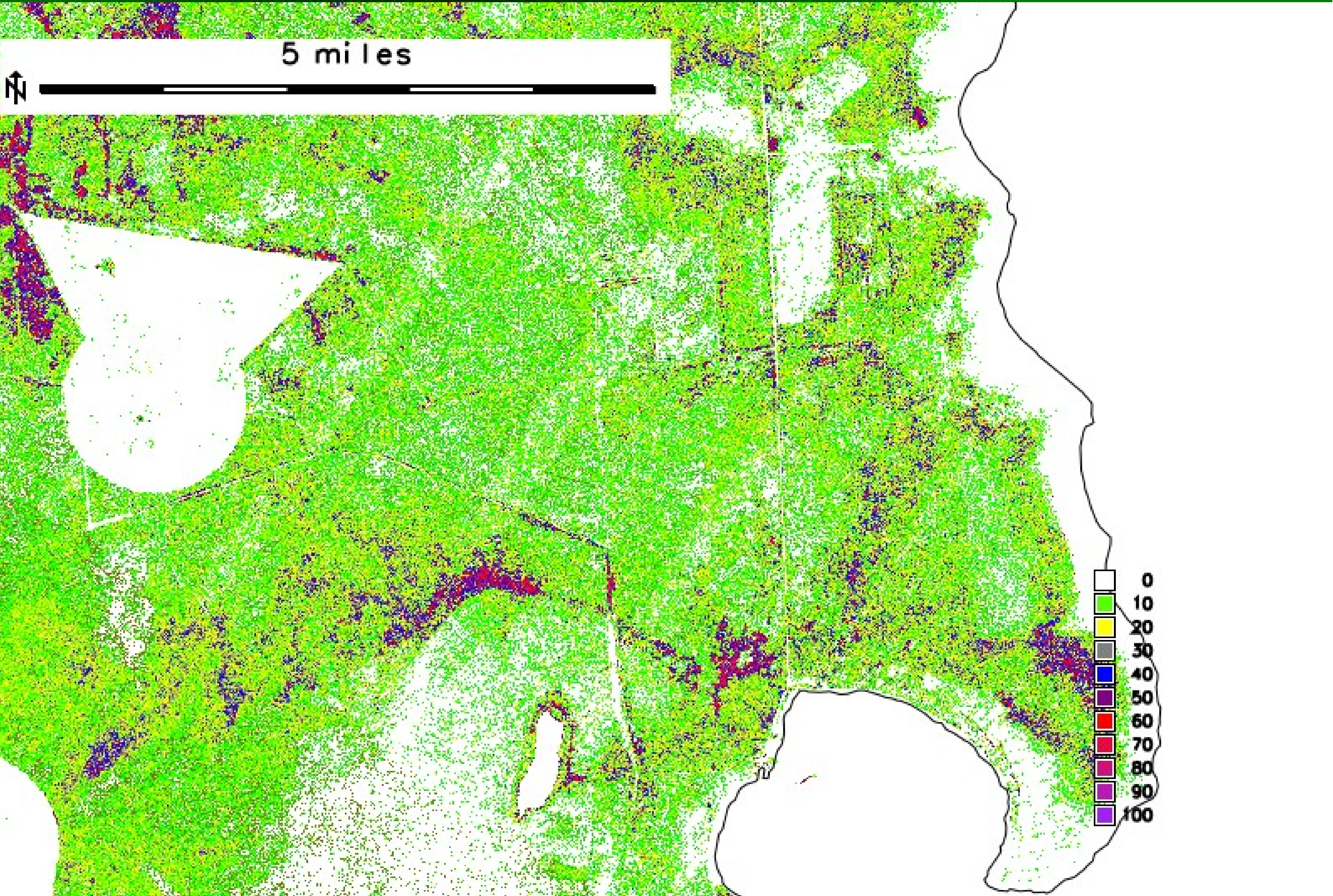
Percent of the Point Cloud in 0 – 10 ft layer



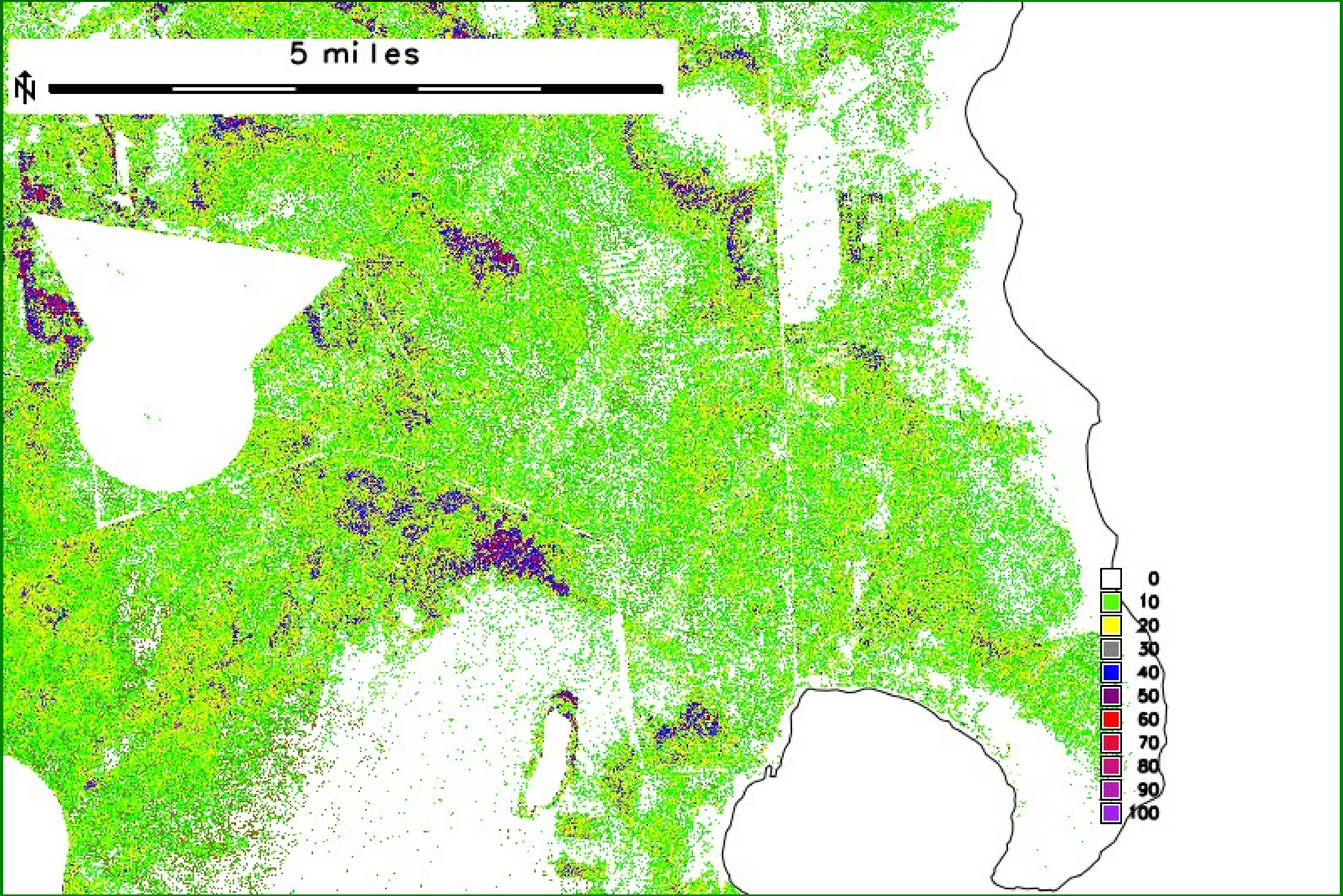
Percent of the Point Cloud in 0 – 10 ft layer



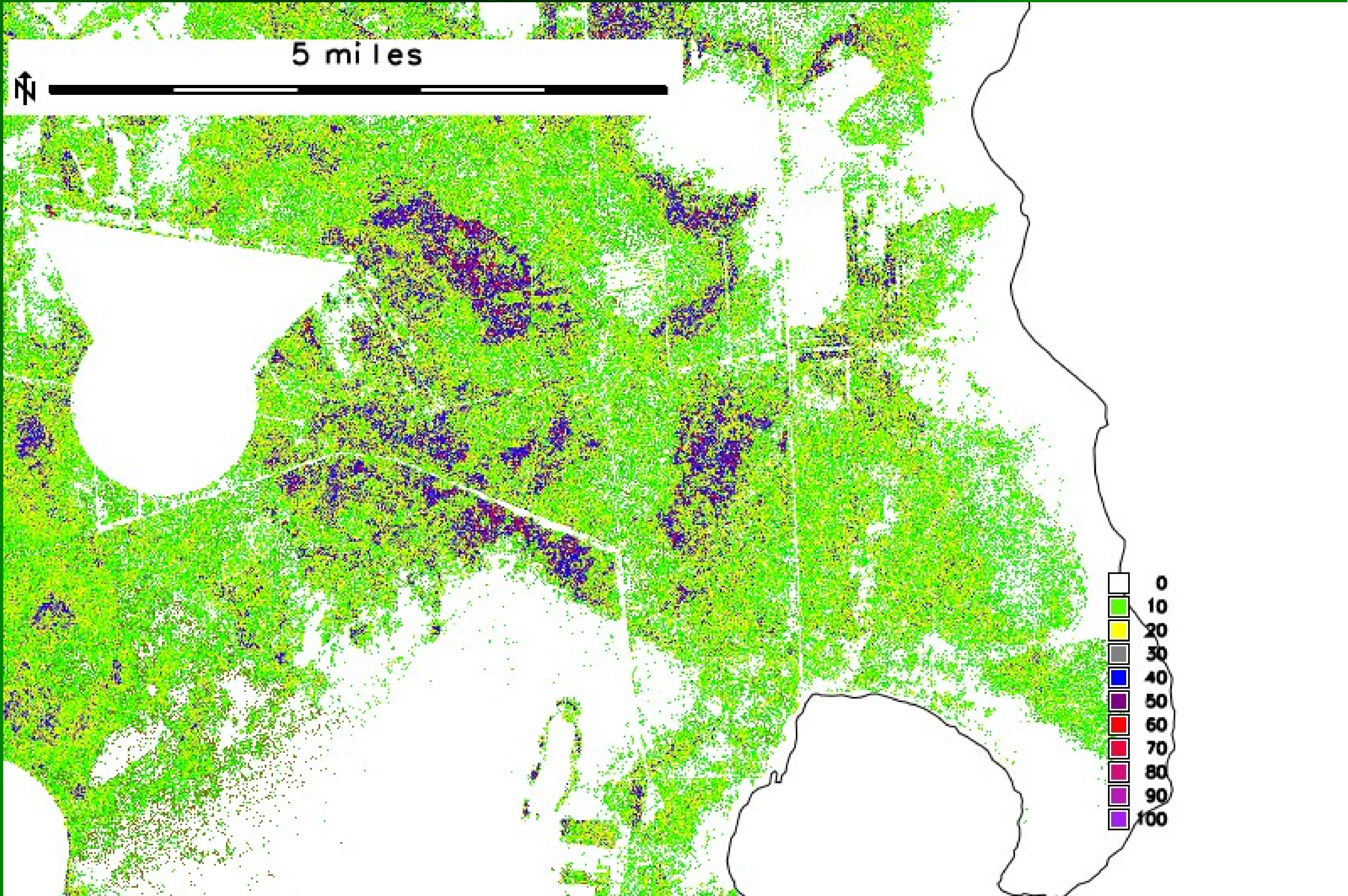
Percent of the Point Cloud in 10 – 20 ft layer



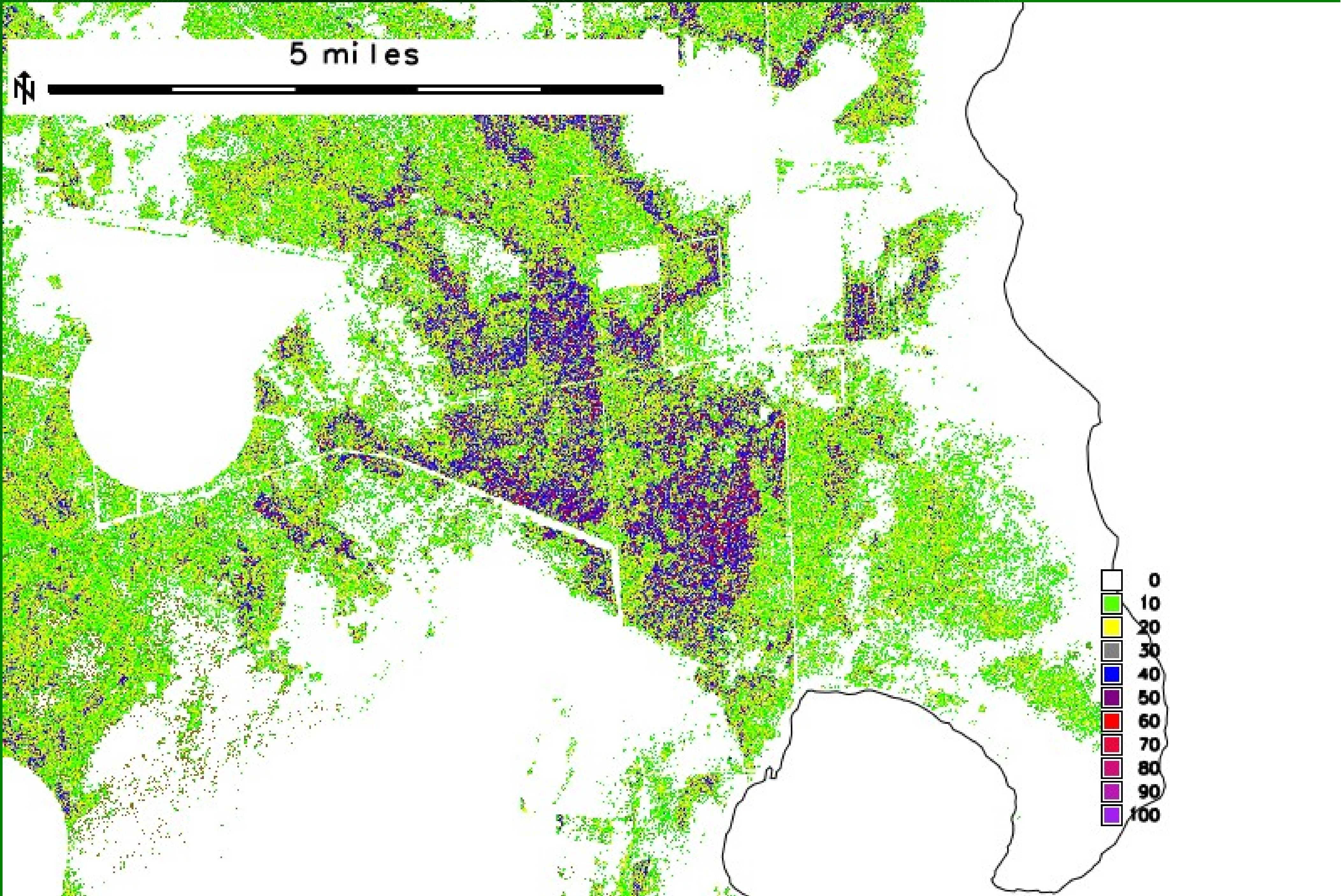
Percent of the Point Cloud in 20 – 30 ft layer



Percent of the Point Cloud in 30 – 40 ft layer



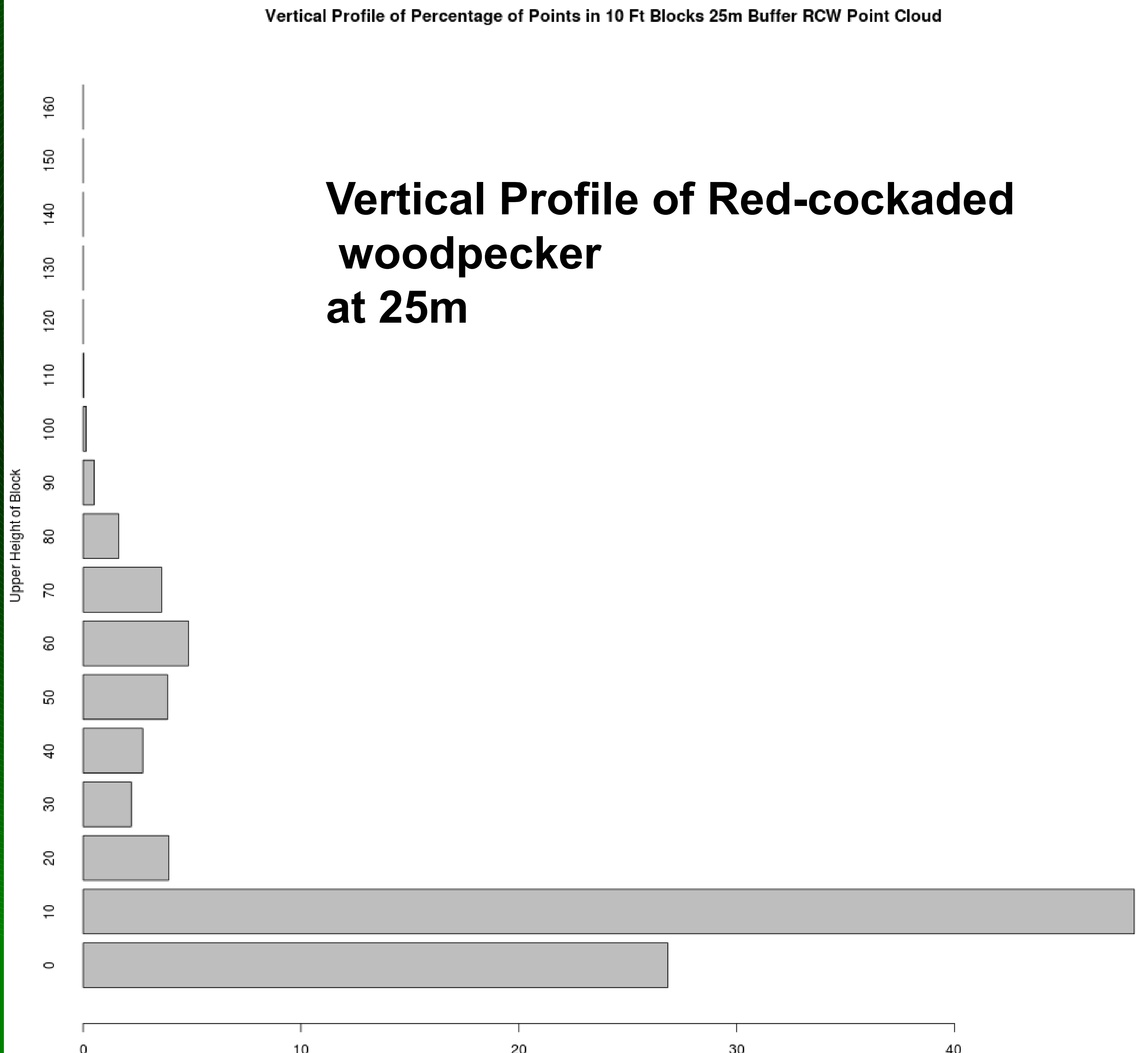
Percent of the Point Cloud in 40 – 50 ft layer



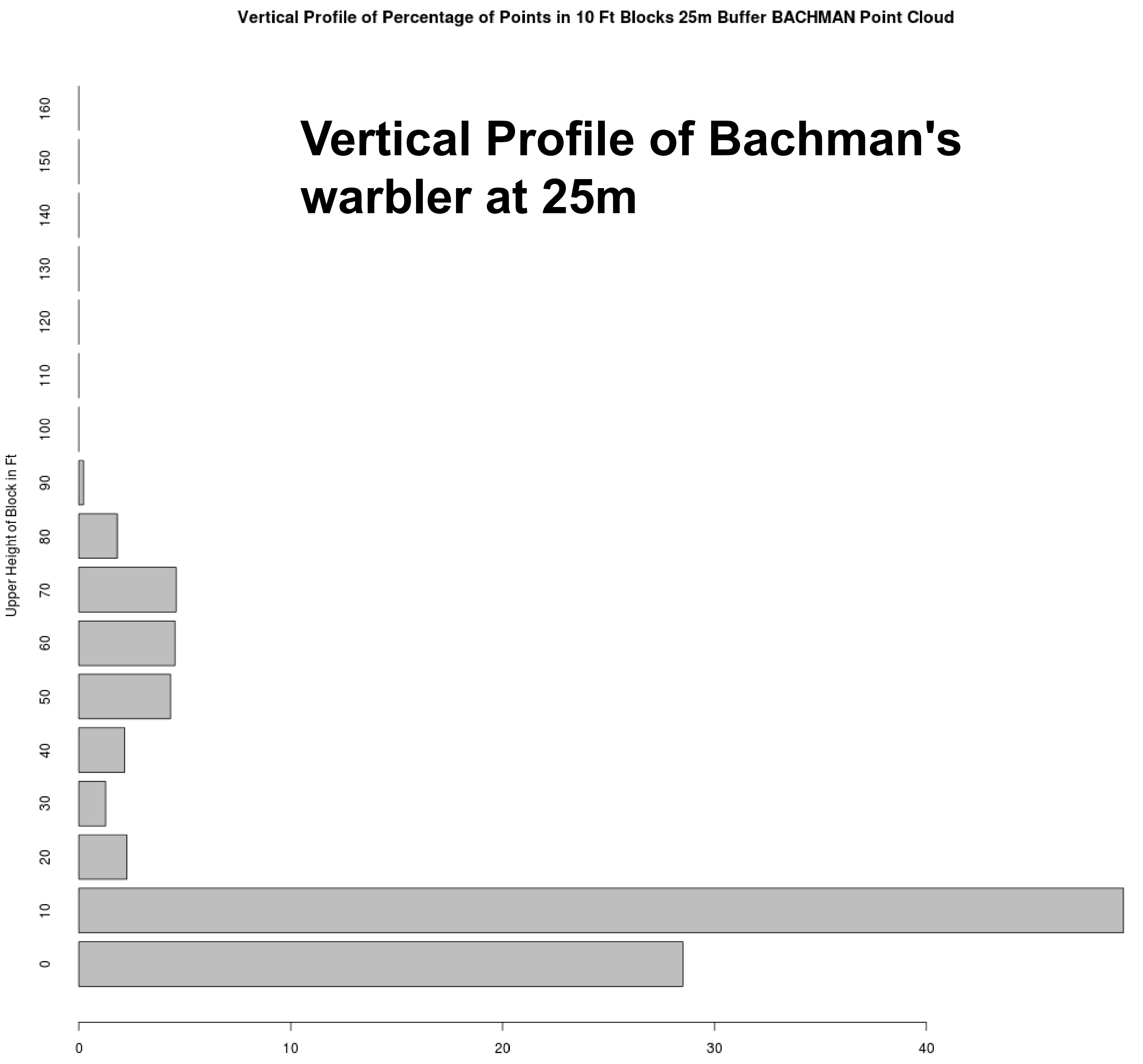
Why Should we care about the percentages by layer of points for each location?

It gives an indication of midstory density , and seems to relate to bird species preferences.

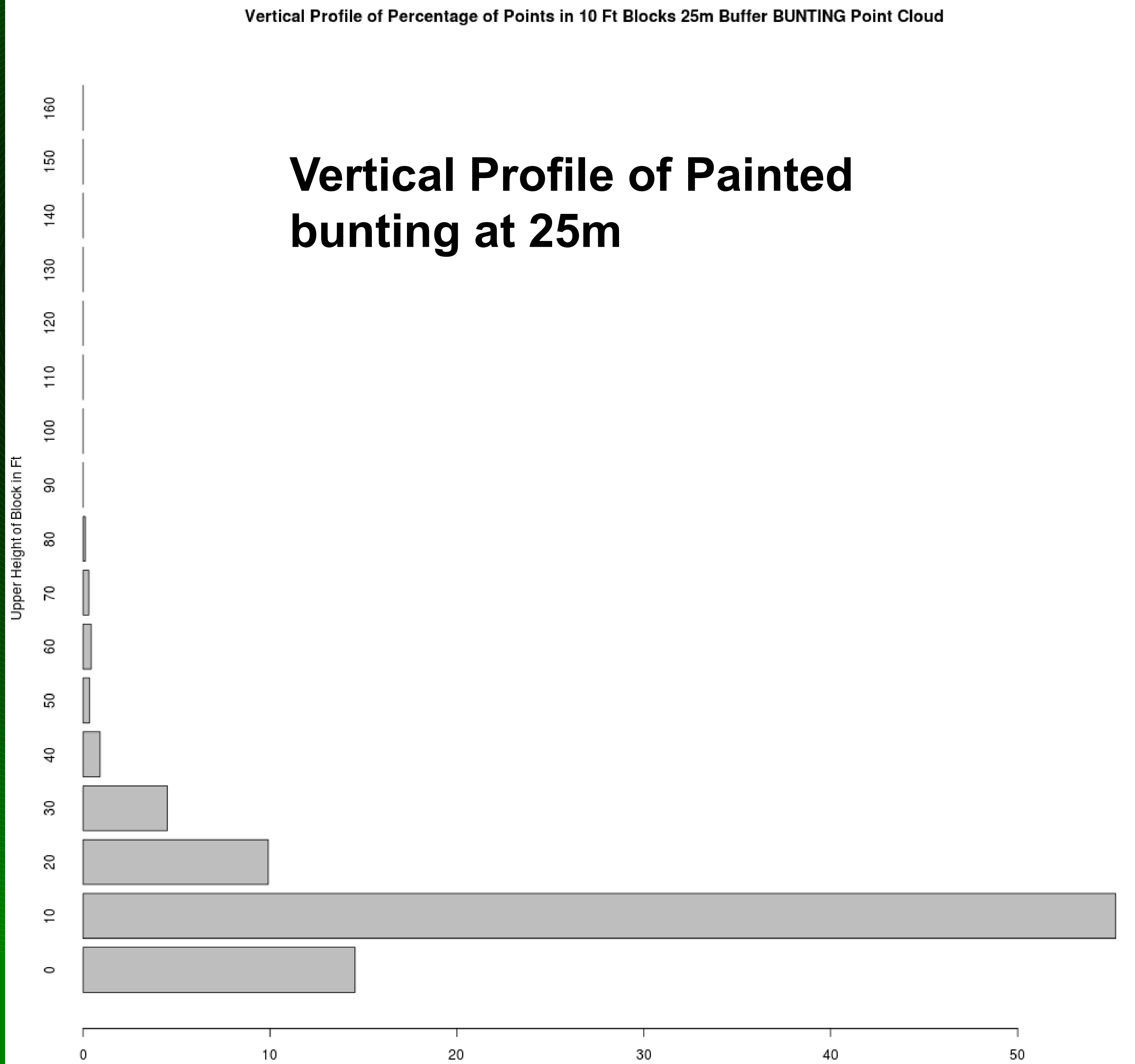
Vertical Profile of Red-cockaded woodpecker at 25m



Vertical Profile of Bachman's warbler at 25m



Vertical Profile of Painted bunting at 25m



This looks interesting, but would need a large parallel processing supercomputer for n dimensional cluster analysis of the different 10 ft layers, along with the other metrics.

This is where Dr. William Hargrove at the Eastern Forest Threat Center, and Forrest Hoffman and Dr. Jitendra Kumar with Oak Ridge National Laboratories stepped into the picture.

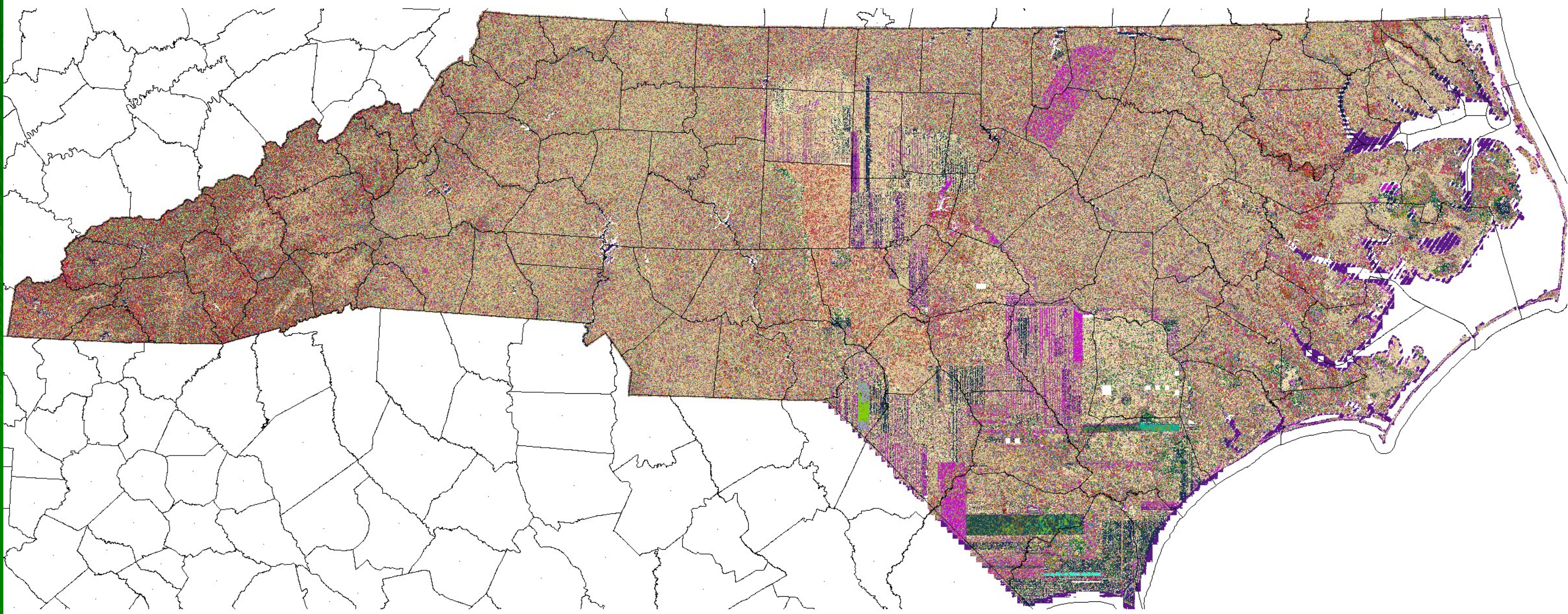
Enter Titan at Oak Ridge National Laboratory:

http://en.wikipedia.org/wiki/Titan_%2

18,688 CPUs paired with an equal number of GPUs

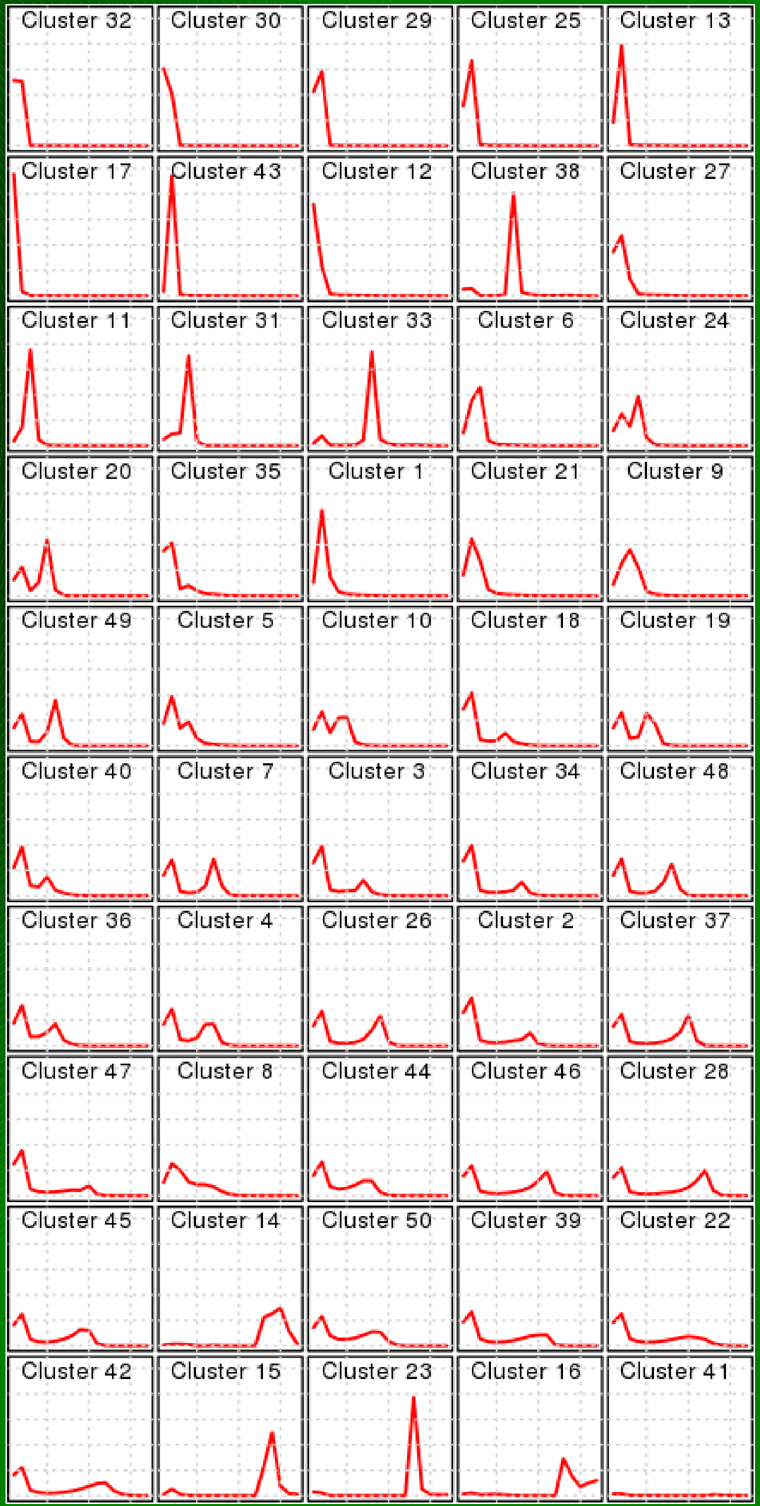
17 layers of the forest structure data was split into a file of 755 million lines with 17 attributes and passed it along for cluster analysis to the Titan Supercomputer to create 50 clusters and 100 clusters

50 Clusters from 17 layers

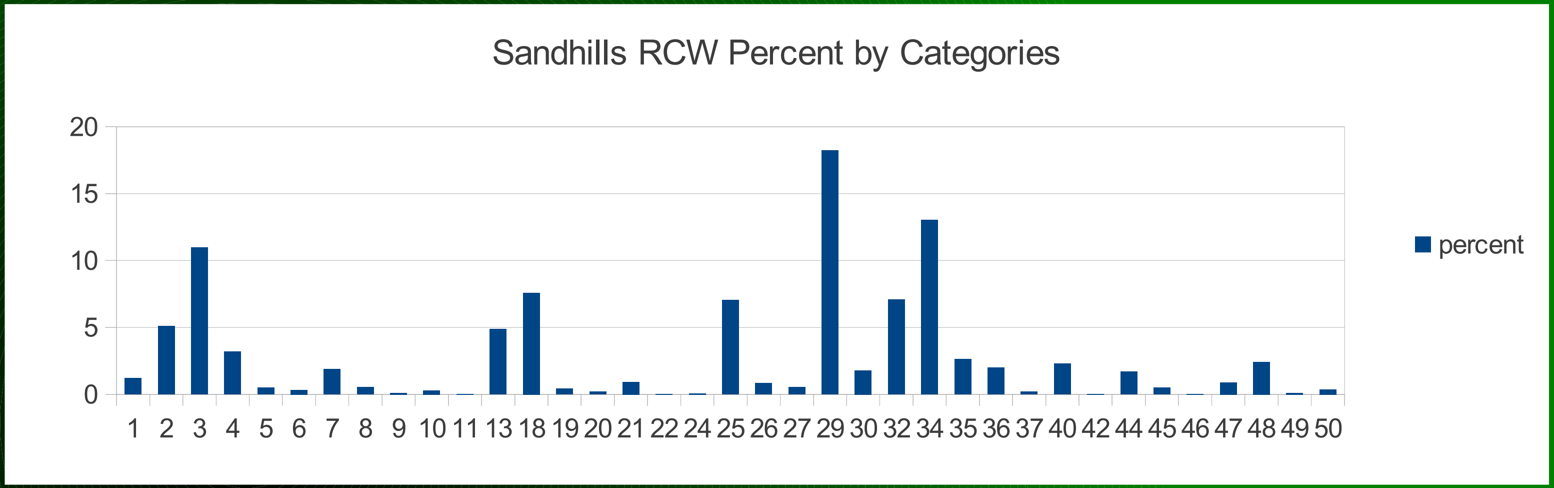


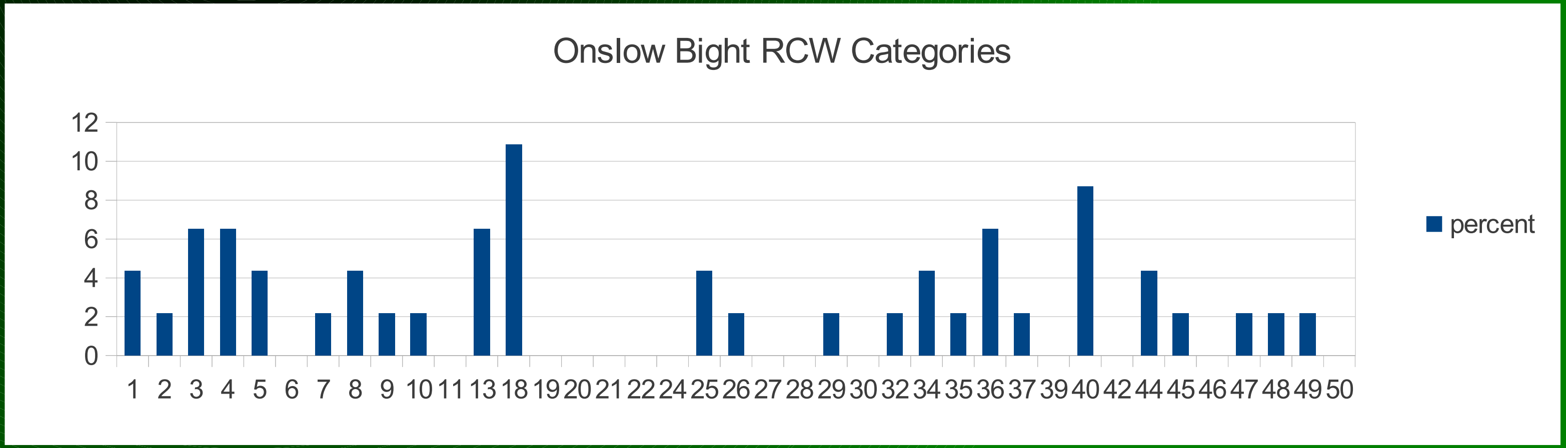
The bad LiDAR data stands out!

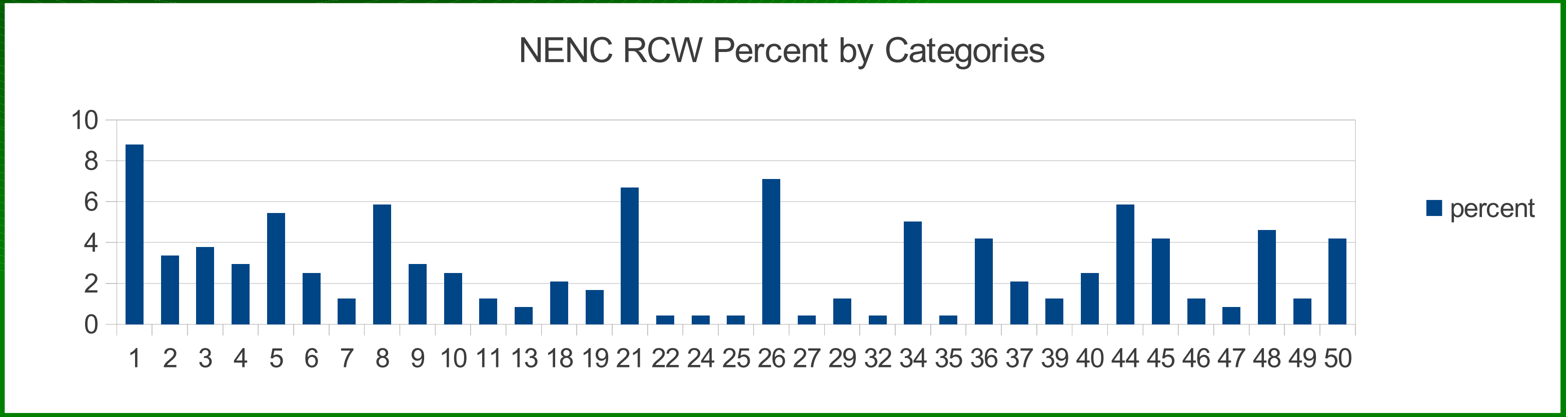
Clustering by similar structures

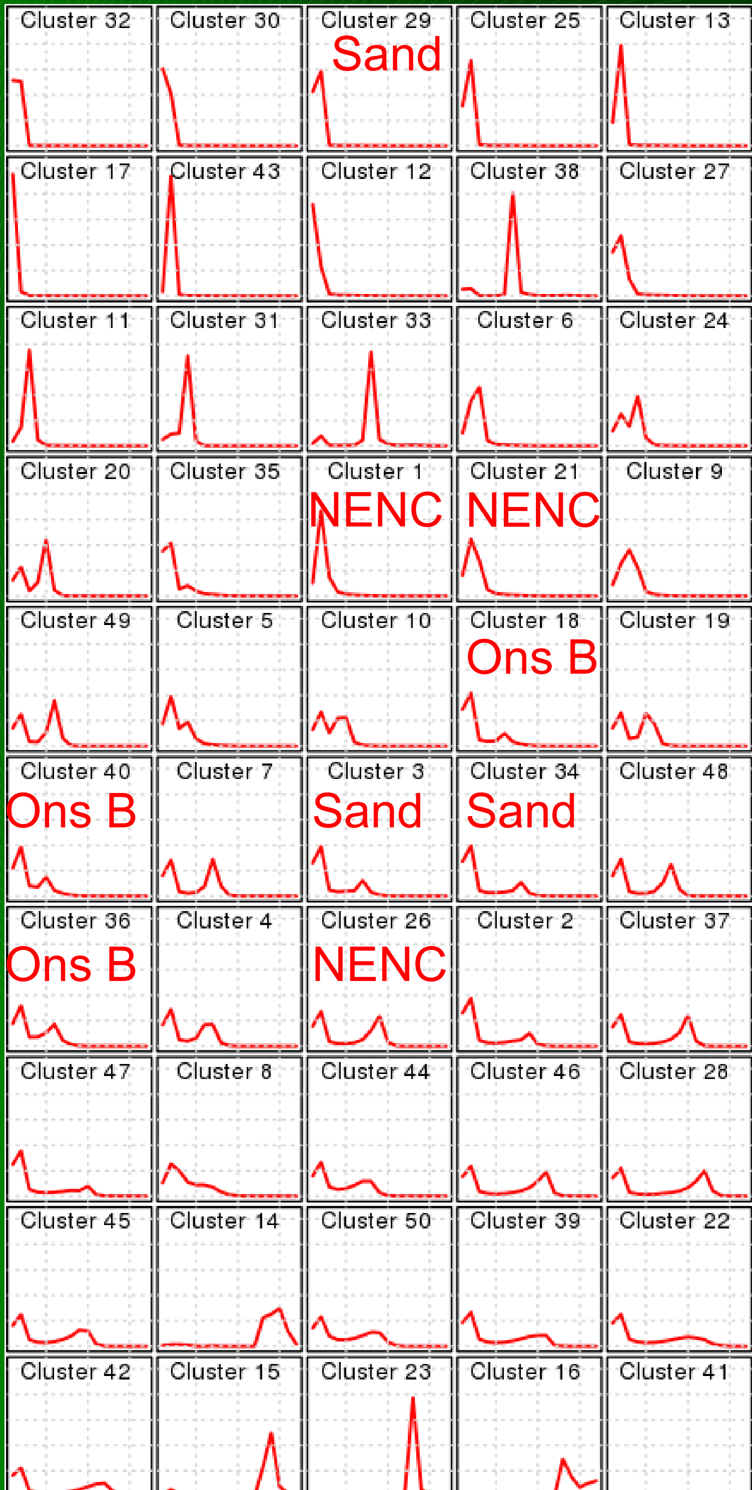


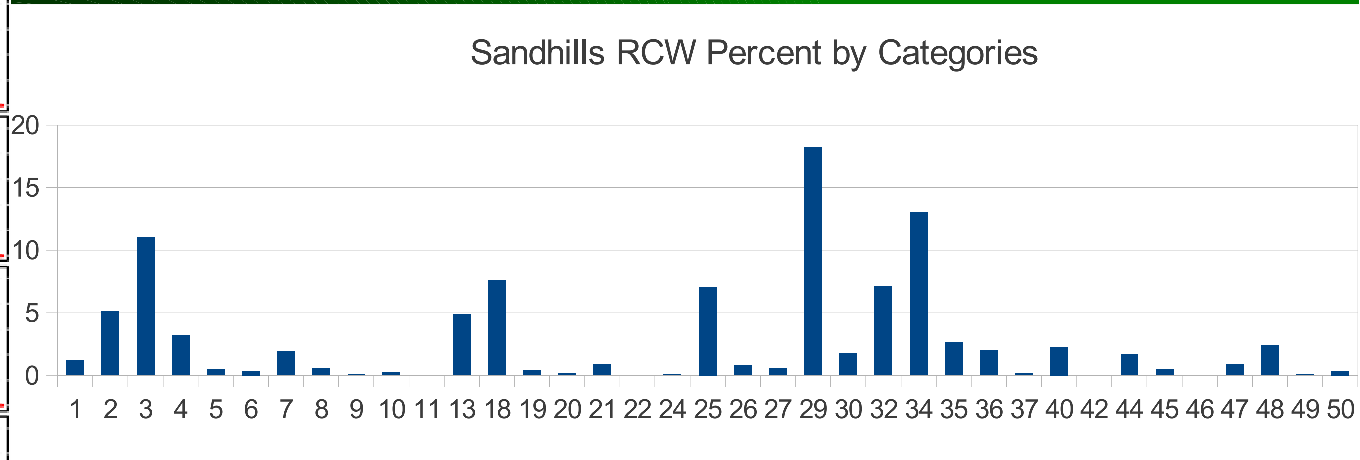
Go back to the RCW data and count the cells in the 25m buffer and collect them by category (excluding the “Bad” Lidar data)

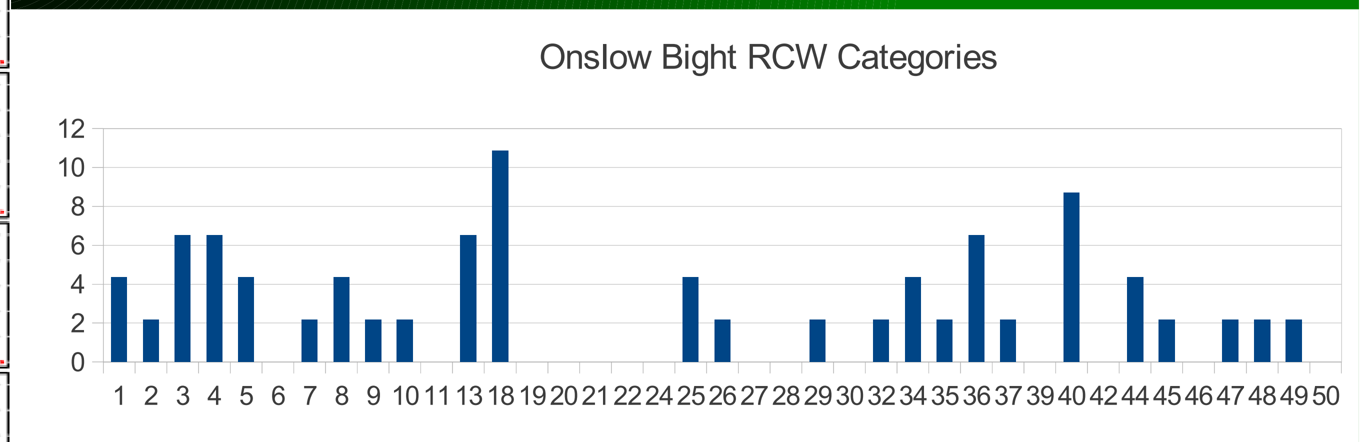

Sandhills RCW Percent by Categories
20
15
10
5
0
1 2 3 4 5 6 7 8 9 10 11 13 18 19 20 21 22 24 25 26 27 29 30 32 34 35 36 37 40 42 44 45 46 47 48 49 50
percent

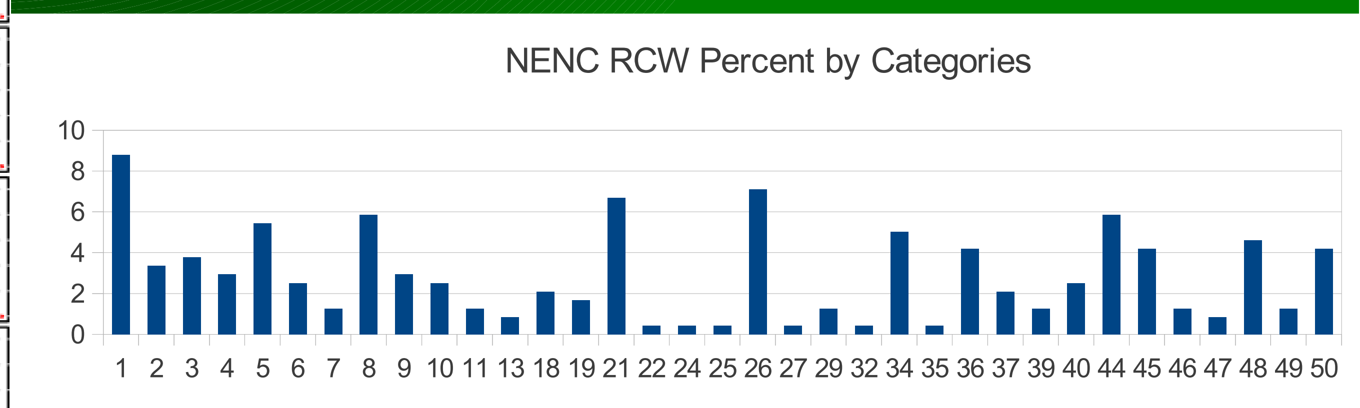

Onslow Bight RCW Categories
12
10
8
6
4
2
0
1 2 3 4 5 6 7 8 9 10 11 13 18 19 20 21 22 24 25 26 27 28 29 30 32 34 35 36 37 39 40 42 44 45 46 47 48 49 50
percent


NENC RCW Percent by Categories
10
8
6
4
2
0
1 2 3 4 5 6 7 8 9 10 11 13 18 19 21 22 24 25 26 27 29 32 34 35 36 37 39 40 44 45 46 47 48 49 50
percent


Cluster 32
Cluster 30
Cluster 29
Sand
Cluster 25
Cluster 13
Cluster 17
Cluster 43
Cluster 12
Cluster 38
Cluster 27
Cluster 11
Cluster 31
Cluster 33
Cluster 6
Cluster 24
Cluster 20
Cluster 35
Cluster 1
NENC
Cluster 21
NENC
Cluster 9
Cluster 49
Cluster 5
Cluster 10
Cluster 18
Ons B
Cluster 19
Cluster 40
Ons B
Cluster 7
Cluster 3
Sand
Cluster 34
Sand
Cluster 48
Cluster 36
Ons B
Cluster 4
Cluster 26
NENC
Cluster 2
Cluster 37
Cluster 47
Cluster 8
Cluster 44
Cluster 46
Cluster 28
Cluster 45
Cluster 14
Cluster 50
Cluster 39
Cluster 22
Cluster 42
Cluster 15
Cluster 23
Cluster 16
Cluster 41


Sandhills RCW Percent by Categories
0
5
10
15
20
1 2 3 4 5 6 7 8 9 10 11 13 18 19 20 21 22 24 25 26 27 29 30 32 34 35 36 37 40 42 44 45 46 47 48 49 50


Onslow Bight RCW Categories
0
2
4
6
8
10
12
1 2 3 4 5 6 7 8 9 10 11 13 18 19 20 21 22 24 25 26 27 28 29 30 32 34 35 36 37 39 40 42 44 45 46 47 48 49 50


NENC RCW Percent by Categories
0
2
4
6
8
10
1 2 3 4 5 6 7 8 9 10 11 13 18 19 21 22 24 25 26 27 29 32 34 35 36 37 39 40 44 45 46 47 48 49 50

Still need to perform analysis on much higher density lidar data that overlap with the original data set to see if similar patterns emerge.

LiDAR - derived Canopy Statistics are statewide data sets and seem to be following known species preferences and giving distinct patterns in the data for each species of bird.

There may be other plant or animal species that show structure patterns as well. The data may also be useful for other uses such as fire fuels estimation.

Questions?