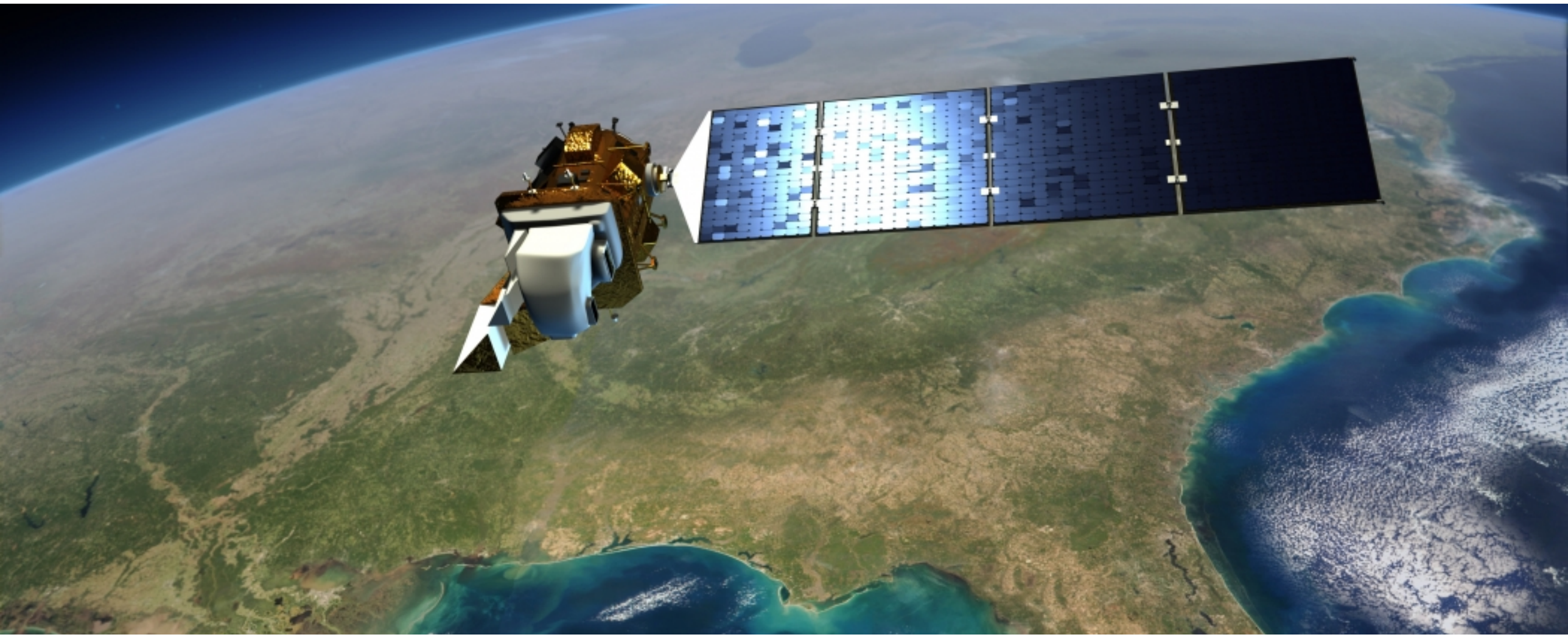
LandSat Image Classification



Steven Bierer

Metis Data Science Bootcamp, Seattle

LandSat Program

- Satellite images of Earth’s surface
- LandSat 4 launched in 1982
- Multiple spectral bands
- Used in agriculture, ecology

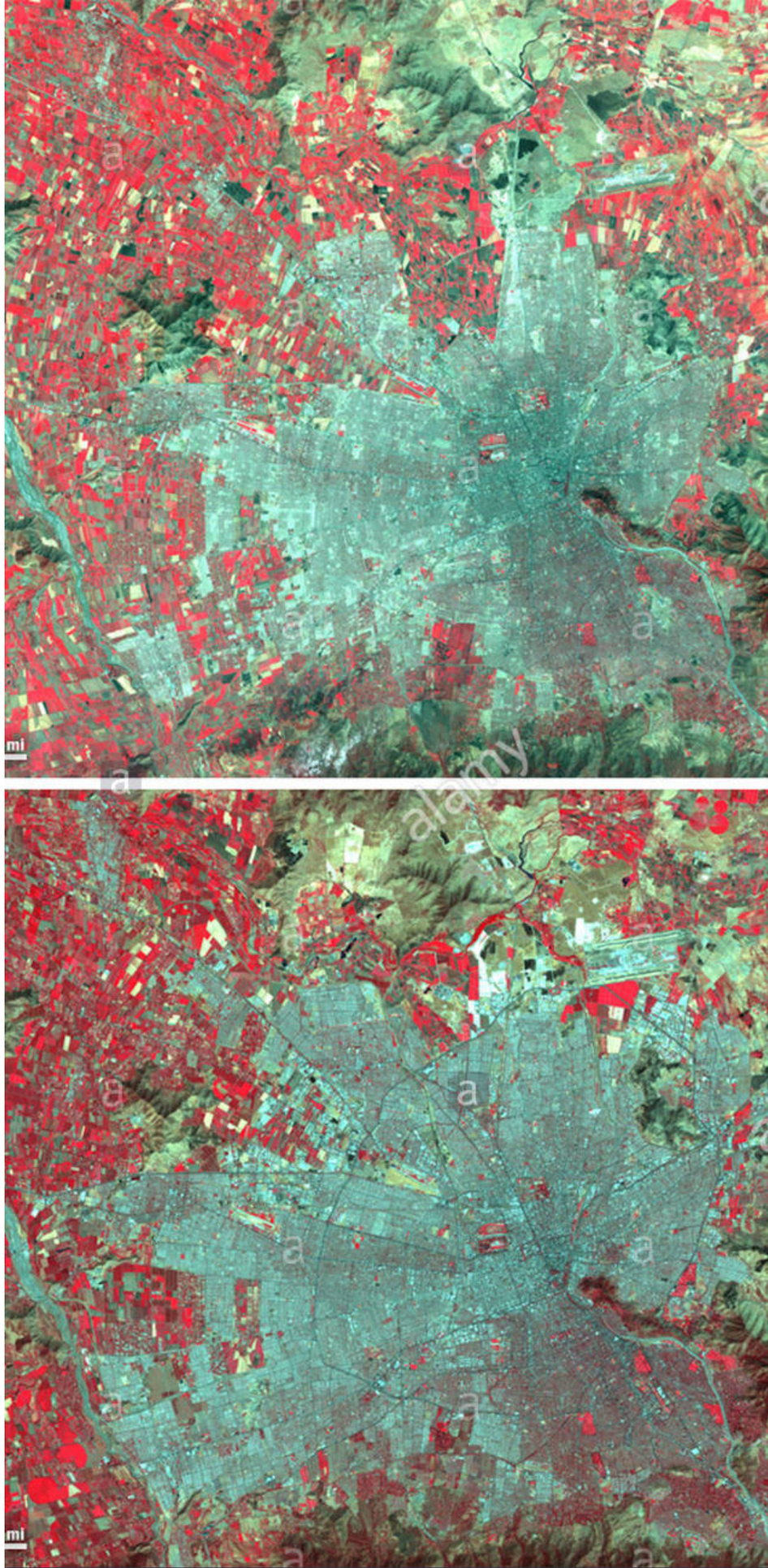
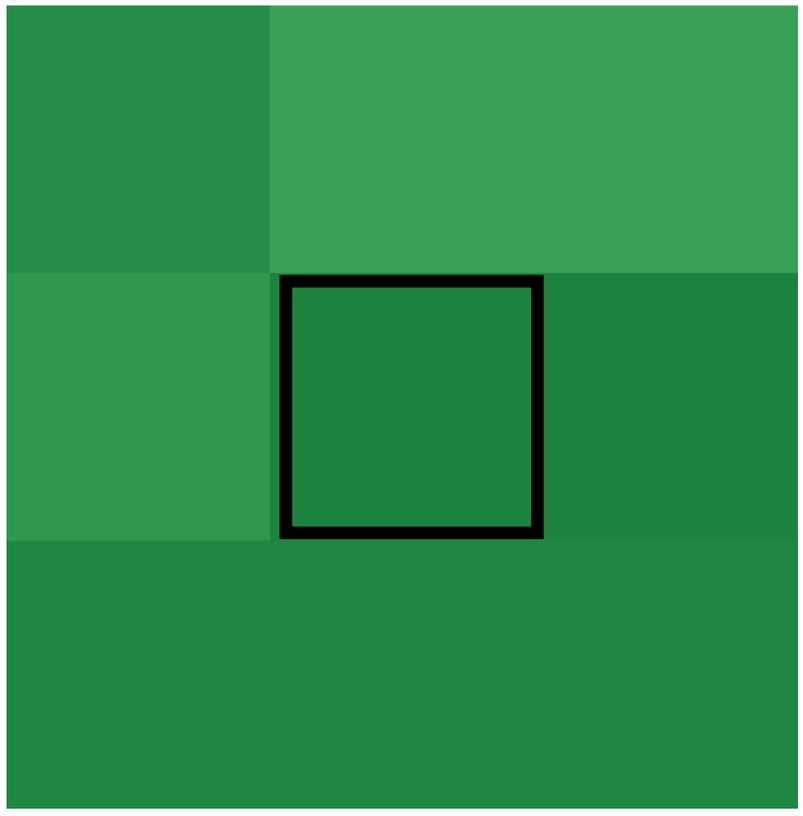


Image data set

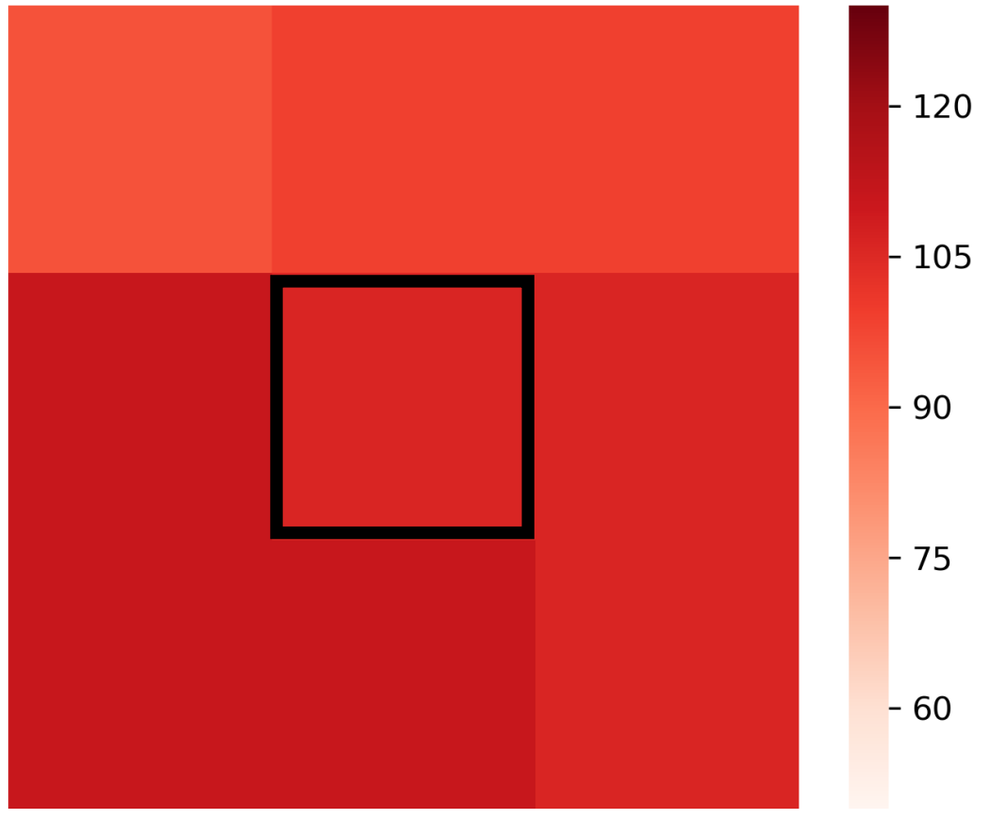
- Source: UC Irvine *Machine Learning Repository*

- 6435 labeled images
  - 3x3 grid of pixels (each 80m x 80m)
- Four spectral components
  - *Green, Red, Near IR, IR*
  - Magnitudes of 0 - 255
- Six labeled terrain types
  - *Cotton Crop, Vegetation, Red Soil, Grey Soil, Damp Soil, Very Damp Soil*

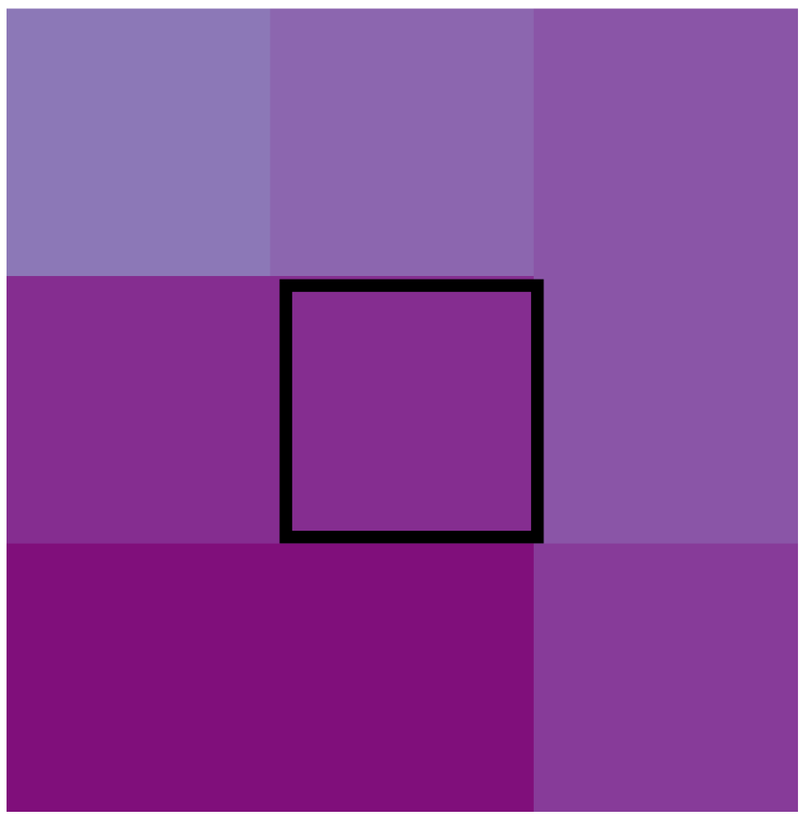
Green Component



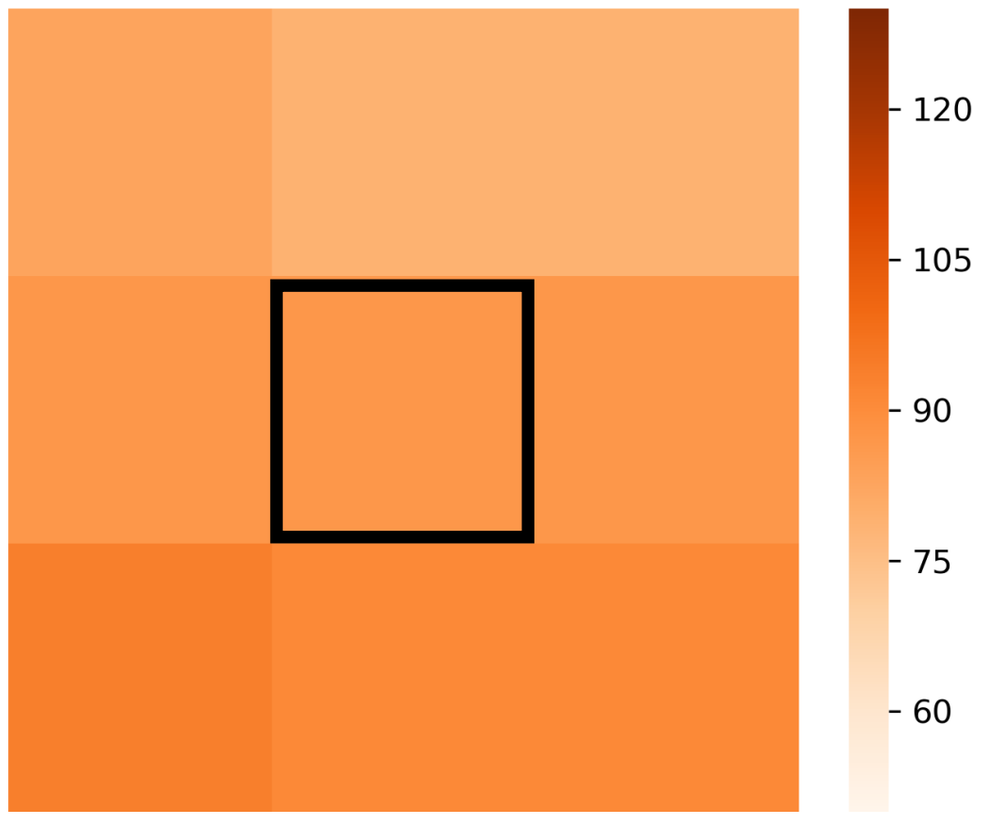
Red Component



Near IR Component

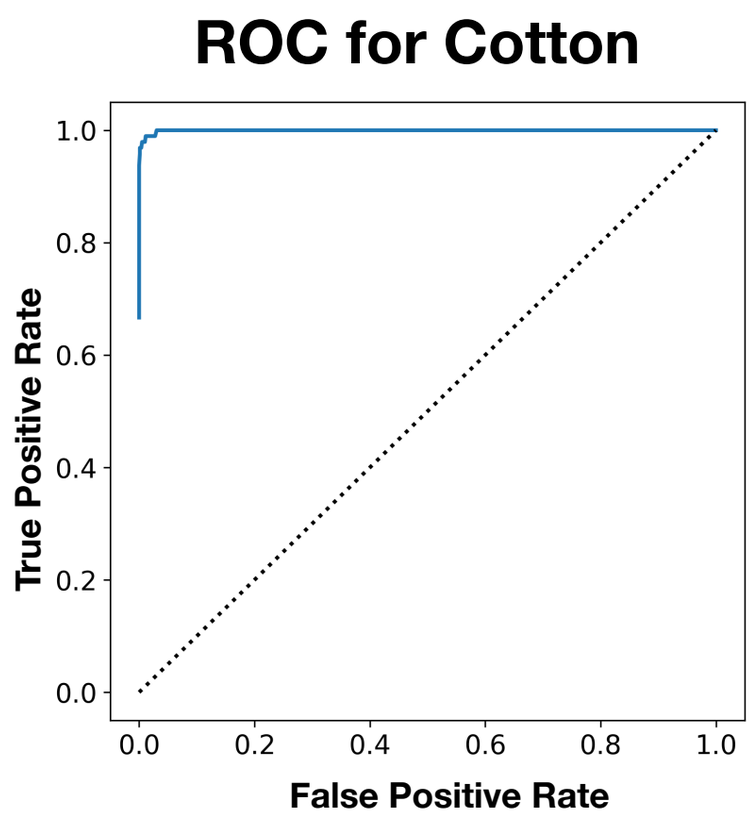


IR Component



Classification training

- Cross-validation scoring based on accuracy
  - Weighted to offset class imbalance
- Principal components
  - 36 -> 7 features



Model	Transform	Parameter	Optimal Val.	Accuracy
KNN	None	# Neighbors	4	0.83
	PC		8	0.79
Logistic Reg.	None	C Value	0.1	0.74
	PC		10	0.73
Random Forest	None	# Estimators	100	0.84
	PC		100	0.83

0.99 !
(full training)

Cost-benefit heuristic

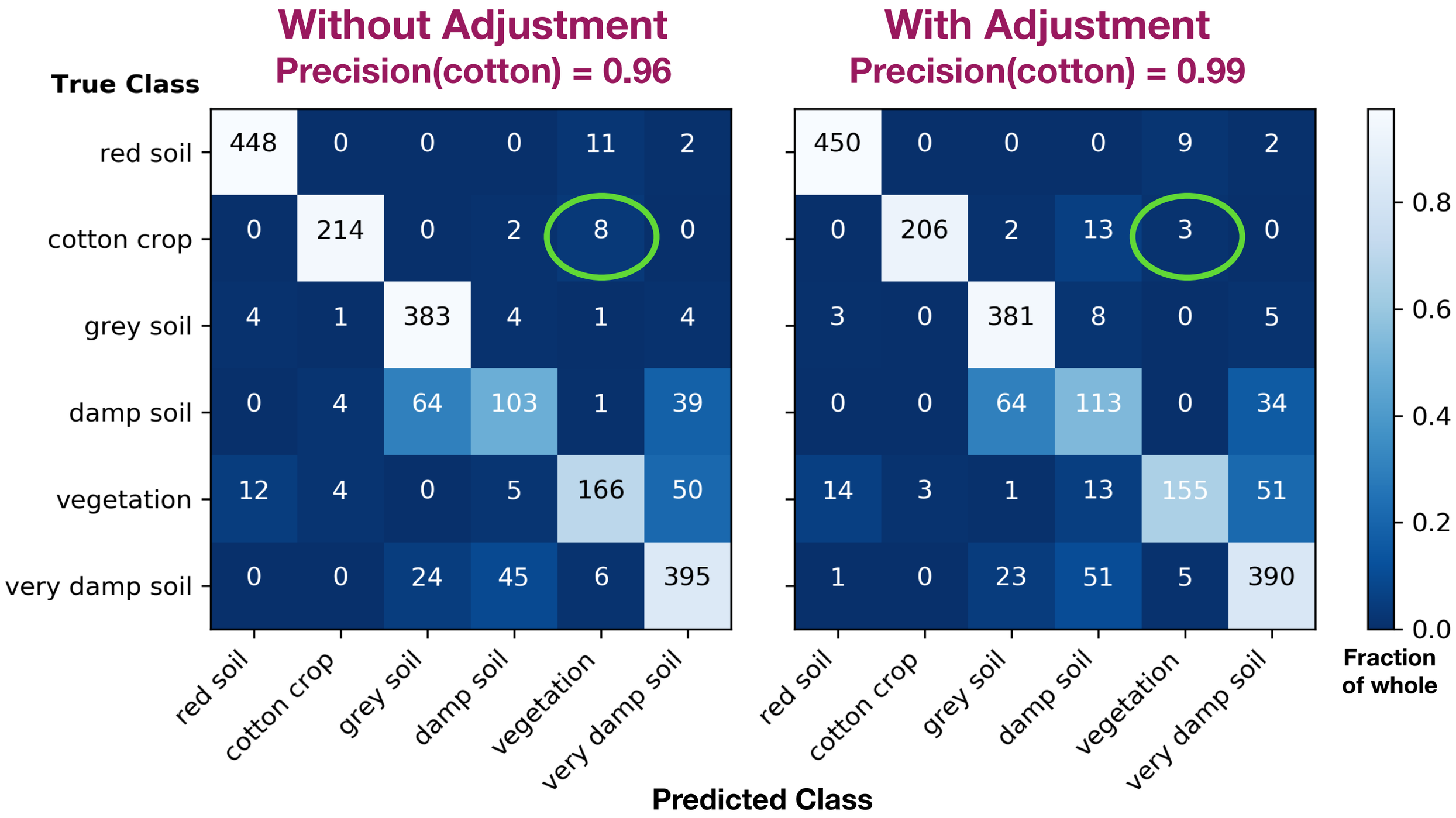
- Extra cost for misclassifying as cotton crop
- Extra cost for misclassifying vegetation with cotton crop
- Don’t care about grey/damp soil confusions

[1-class probability] **[COSTS]** **= [cost score]**

[class probability] **[BENEFITS]** **= [benefit score]**

[class prediction] = argmax(benefit score - cost score)

Random Forest Results



Future considerations

- Rotation-invariant PCA
  - "Hu" Image Moments
- Larger image training set
- More sophisticated classifier
  - e.g. Deep Learning

Appendix

Statistics with cost-benefit adjustment

	precision	recall	f1-score	support
red soil	0.96	0.98	0.97	461.0
cotton crop	0.99	0.92	0.95	224.0
grey soil	0.81	0.96	0.88	397.0
damp soil	0.57	0.54	0.55	211.0
vegetation	0.90	0.65	0.76	237.0
very damp soil	0.81	0.83	0.82	470.0
avg / total	0.85	0.85	0.85	2000.0

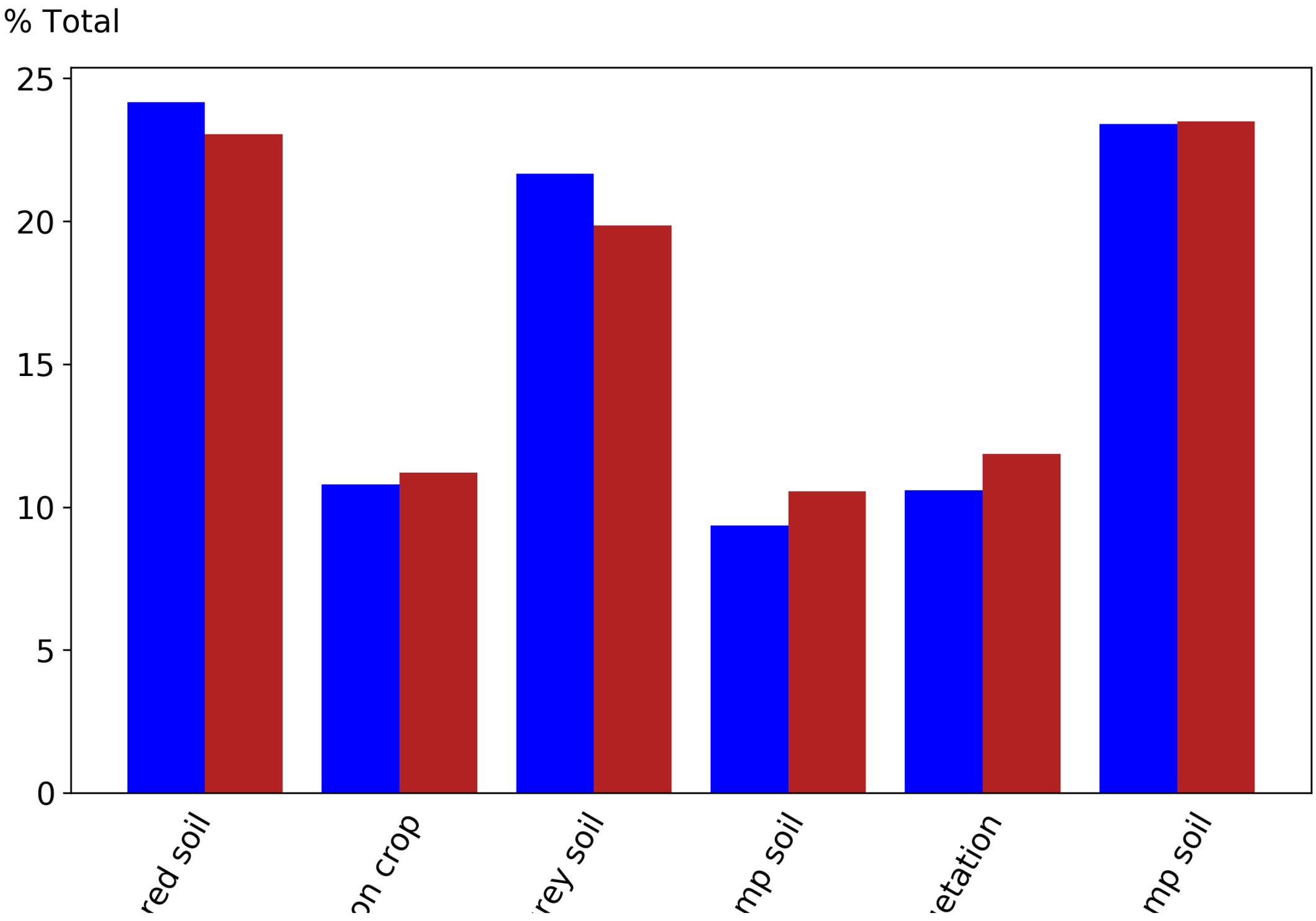
```
-- Metric ----- Average Scores ----
 Accuracy:       0.848
 Precision:      0.840
 Recall:         0.813
 F1 Score:       0.822
```

Statistics without cost-benefit adjustment

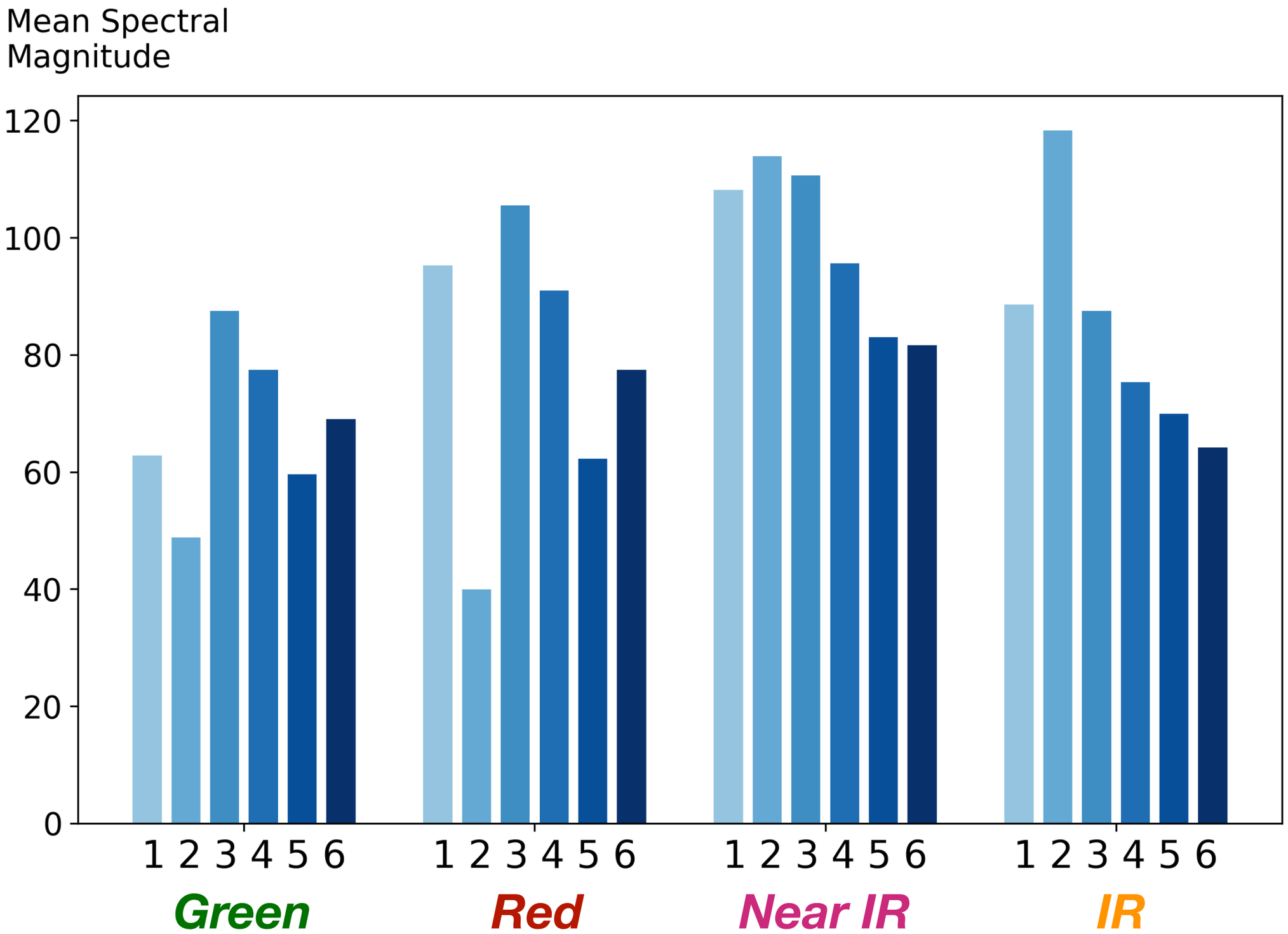
	precision	recall	f1-score	support
red soil	0.97	0.97	0.97	461.0
cotton crop	0.96	0.96	0.96	224.0
grey soil	0.81	0.96	0.88	397.0
damp soil	0.65	0.49	0.56	211.0
vegetation	0.86	0.70	0.77	237.0
very damp soil	0.81	0.84	0.82	470.0
avg / total	0.85	0.85	0.85	2000.0

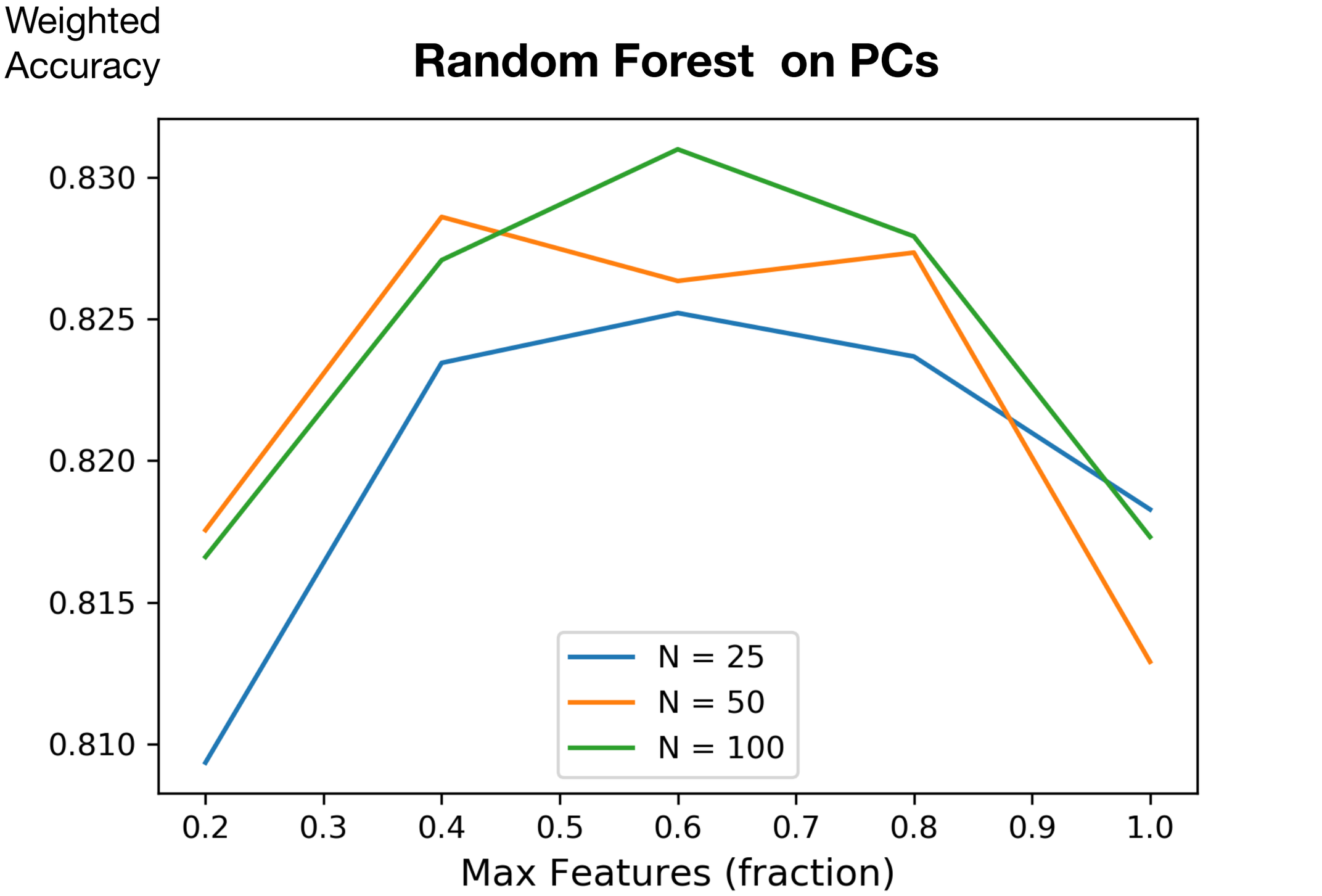
```
-- Metric ----- Average Scores ----
 Accuracy:       0.855
 Precision:      0.842
 Recall:         0.820
 F1 Score:       0.827
```

Training (blue) and test (red) class distributions

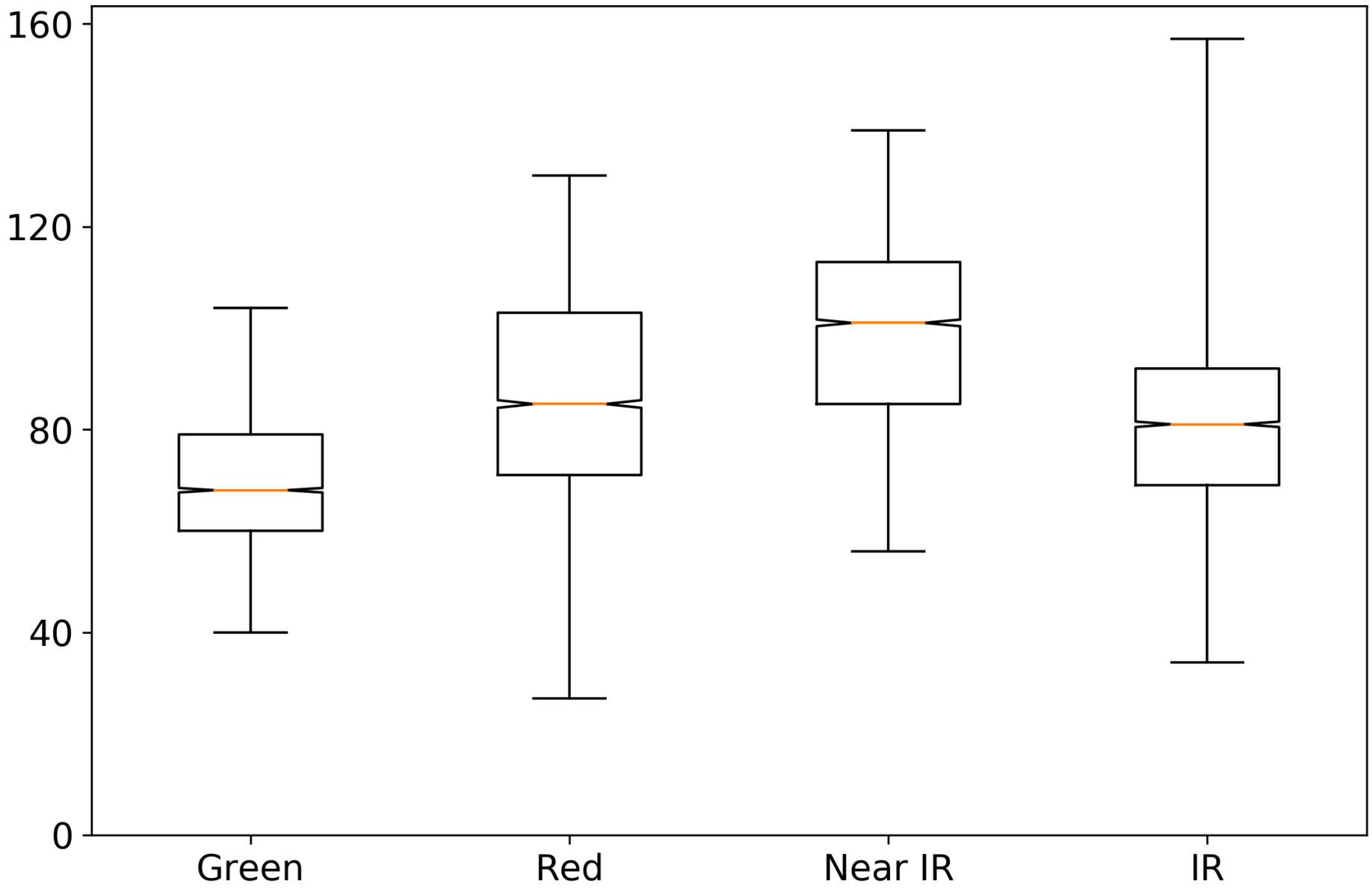


Classify based on “spectral signatures”

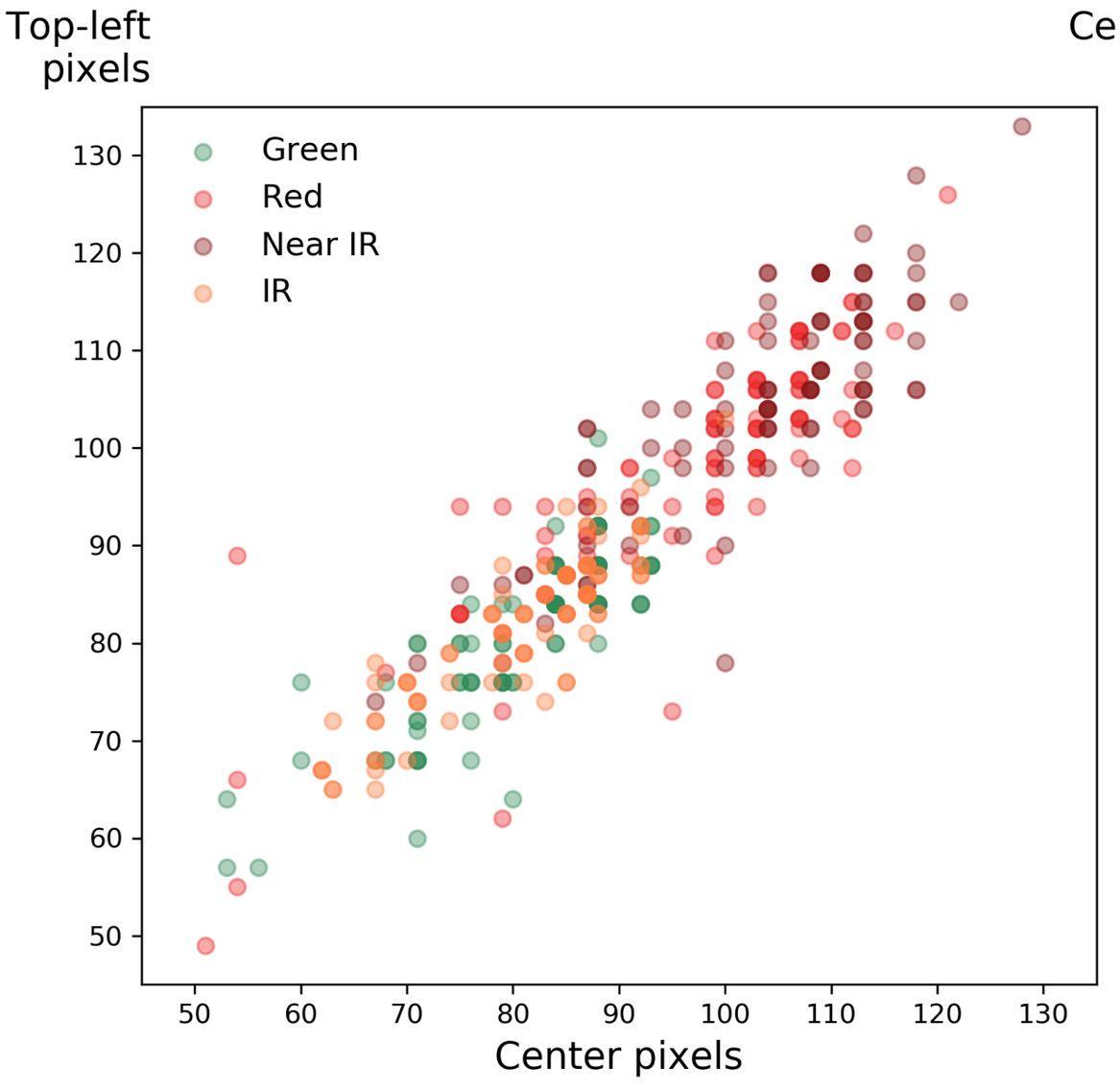


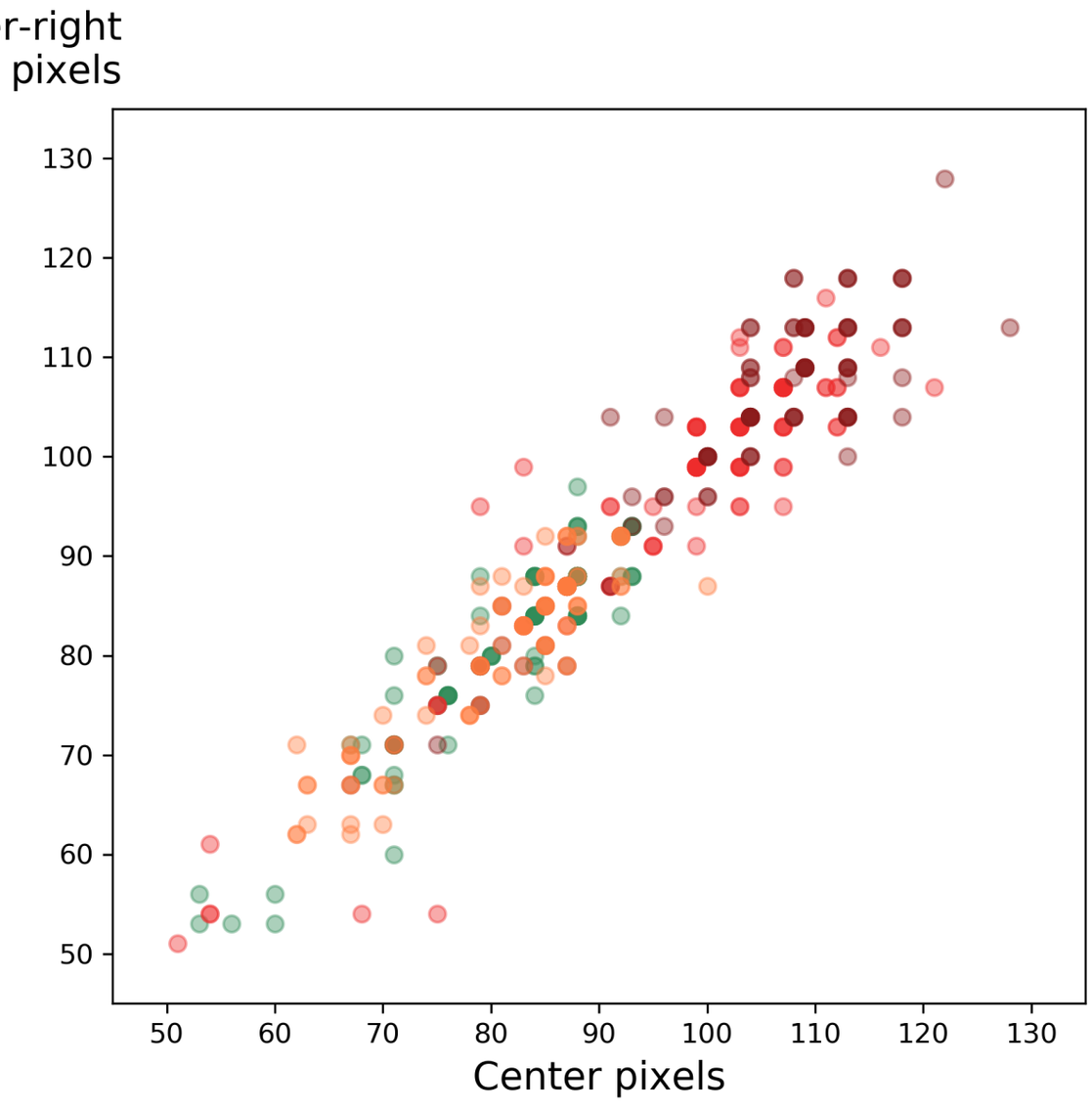

Weighted
Accuracy
Random Forest on PCs
0.830
0.825
0.820
0.815
0.810
N = 25
N = 50
N = 100
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Max Features (fraction)

Spectral Magnitude Distributions



Correlation of spectral magnitudes at two pixels





```
COST = np.eye(6);
COST[1,1]=3;                    # not-p-cotton gives extra penalty to "cotton"
BENEFIT = np.eye(6);
BENEFIT[1,4]=-2;                  # p-cotton suppresses "vegetation"
BENEFIT[2,3]=0.3; BENEFIT[2,5]=0.2  # for p-greysoil, allow confusions with other "grey soil" types
BENEFIT[3,2]=0.3; BENEFIT[3,5]=0.2; # for p-dampsoil, allow confusions with other "grey soil" types
BENEFIT[5,2]=0.2; BENEFIT[5,3]=0.3; # for p-verygreysoil, allow confusions with other "grey soil" types

print(BENEFIT, '\n\n', COST)
```

```
[[ 1.   0.   0.   0.   0.   0. ]
 [ 0.   1.   0.   0.  -2.   0. ]
 [ 0.   0.   1.   0.3  0.   0.2]
 [ 0.   0.   0.3  1.   0.   0.2]
 [ 0.   0.   0.   0.   1.   0. ]
 [ 0.   0.   0.2  0.3  0.   1. ]]

 [[ 1.  0.  0.  0.  0.  0.]
 [ 0.  3.  0.  0.  0.  0.]
 [ 0.  0.  1.  0.  0.  0.]
 [ 0.  0.  0.  1.  0.  0.]
 [ 0.  0.  0.  0.  1.  0.]
 [ 0.  0.  0.  0.  0.  1.]]
```