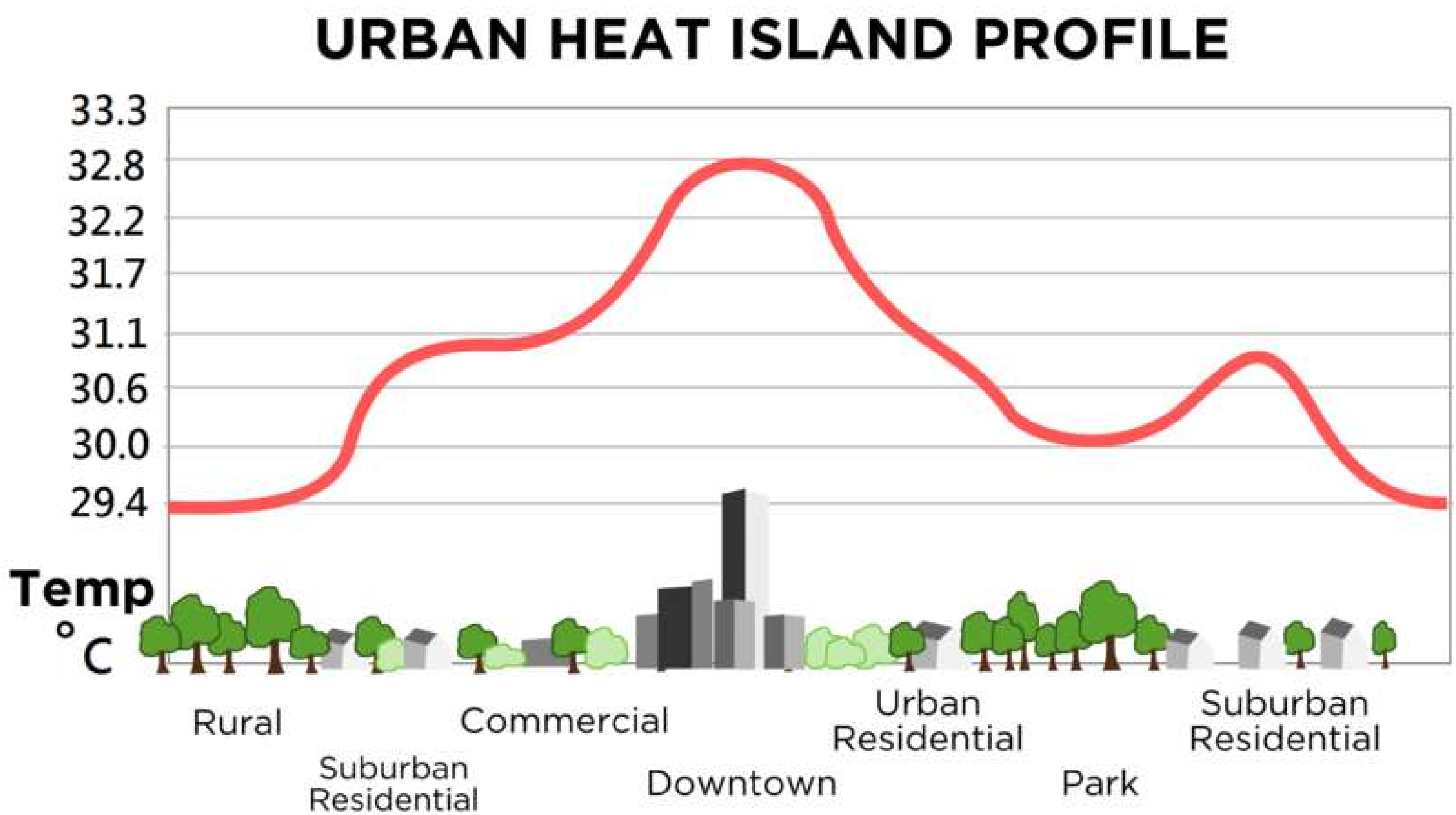
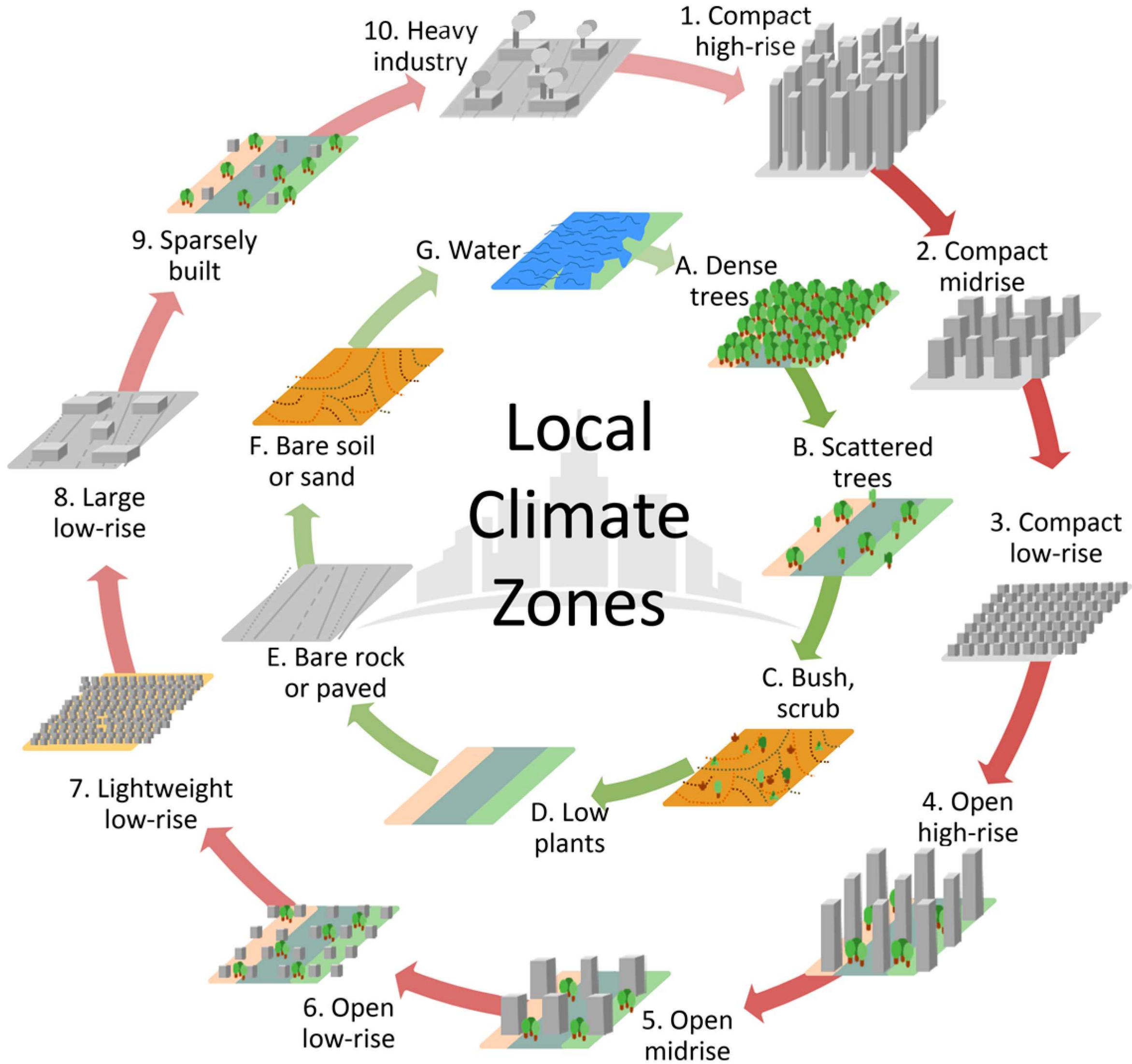
Local Climate Zone Classification Using Random Forests

Ericka B. Smith

3/9/2021


URBAN HEAT ISLAND PROFILE
33.3
32.8
32.2
31.7
31.1
30.6
30.0
29.4
Temp °C
Rural
Suburban Residential
Commercial
Downtown
Urban Residential
Park
Suburban Residential


Local
Climate
Zones
1. Compact
high-rise
2. Compact
midrise
3. Compact
low-rise
4. Open
high-rise
5. Open
midrise
6. Open
low-rise
7. Lightweight
low-rise
8. Large
low-rise
9. Sparsely
built
10. Heavy
industry
A. Dense
trees
B. Scattered
trees
C. Bush,
scrub
D. Low
plants
E. Bare rock
or paved
F. Bare soil
or sand
G. Water

Objective

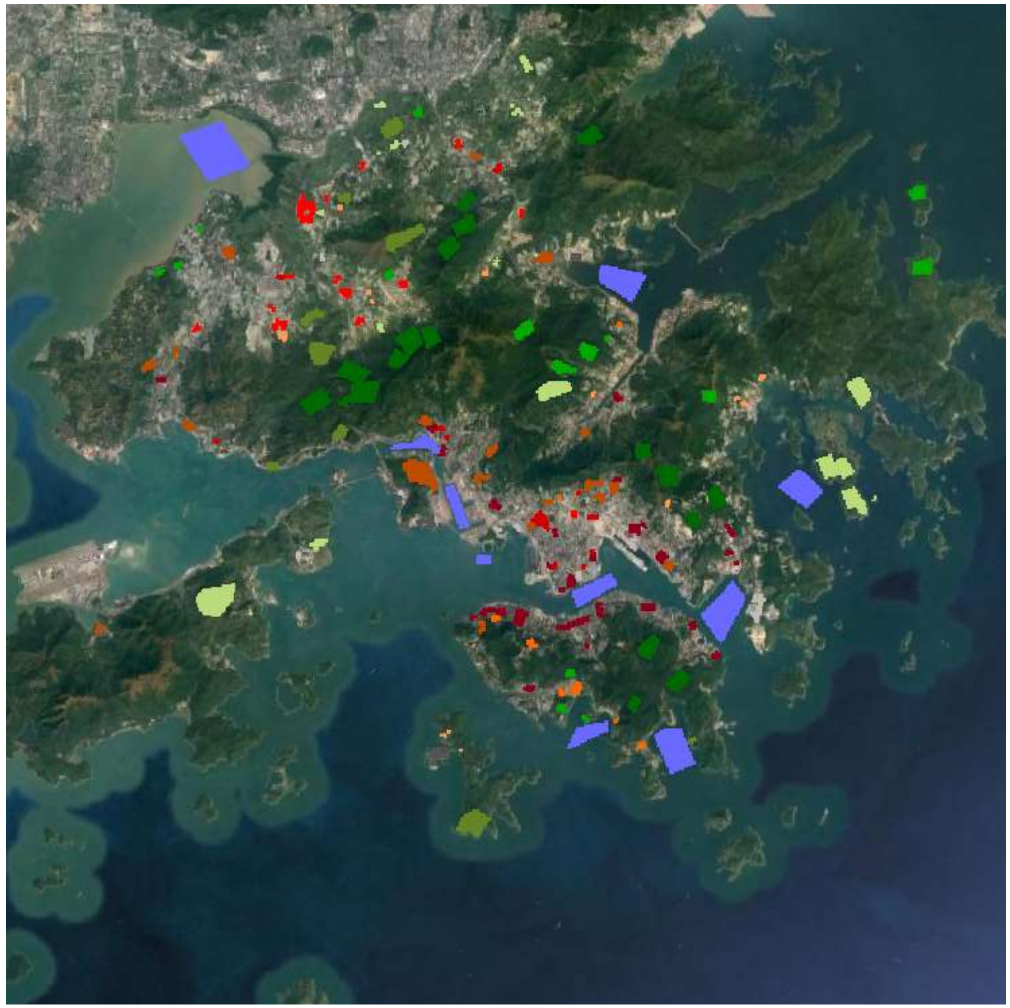
Inspiration:

- *Comparison between convolutional neural networks and random forest for local climate zone classification in mega urban areas using Landsat Images* (Yoo et al., 2019)

My Focus:

- Hong Kong
- Random Forests
- Varying the Number of Trees

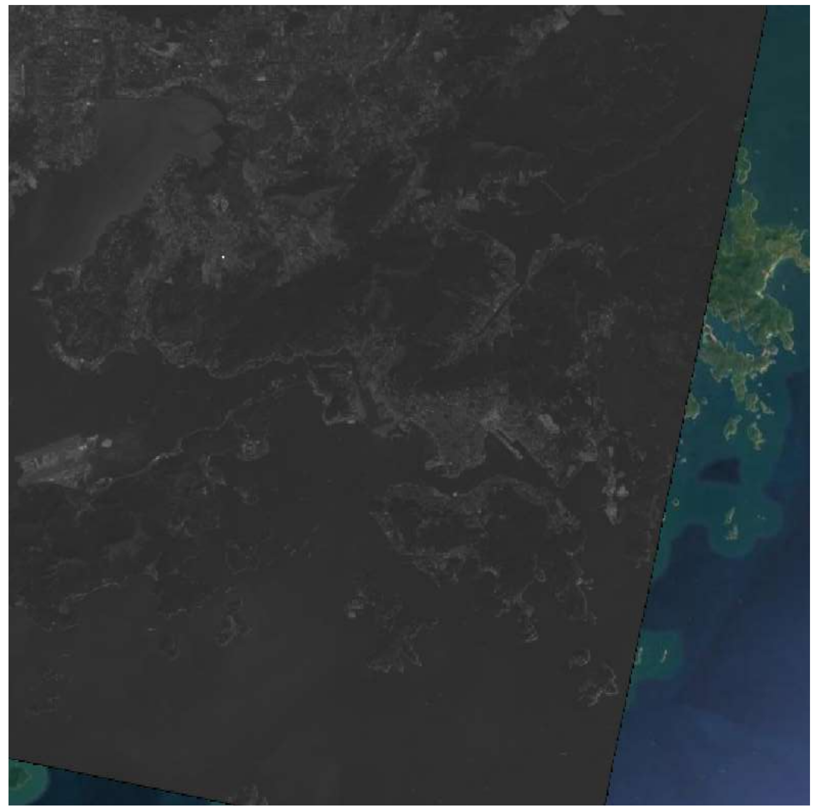
The LCZ reference data



The Landsat 8 data

Acquisition Dates of Each Landsat 8 Scene

Scene	Date
1	29-Nov-2013
2	15-Oct-2014
3	16-Nov-2014
4	18-Oct-2015

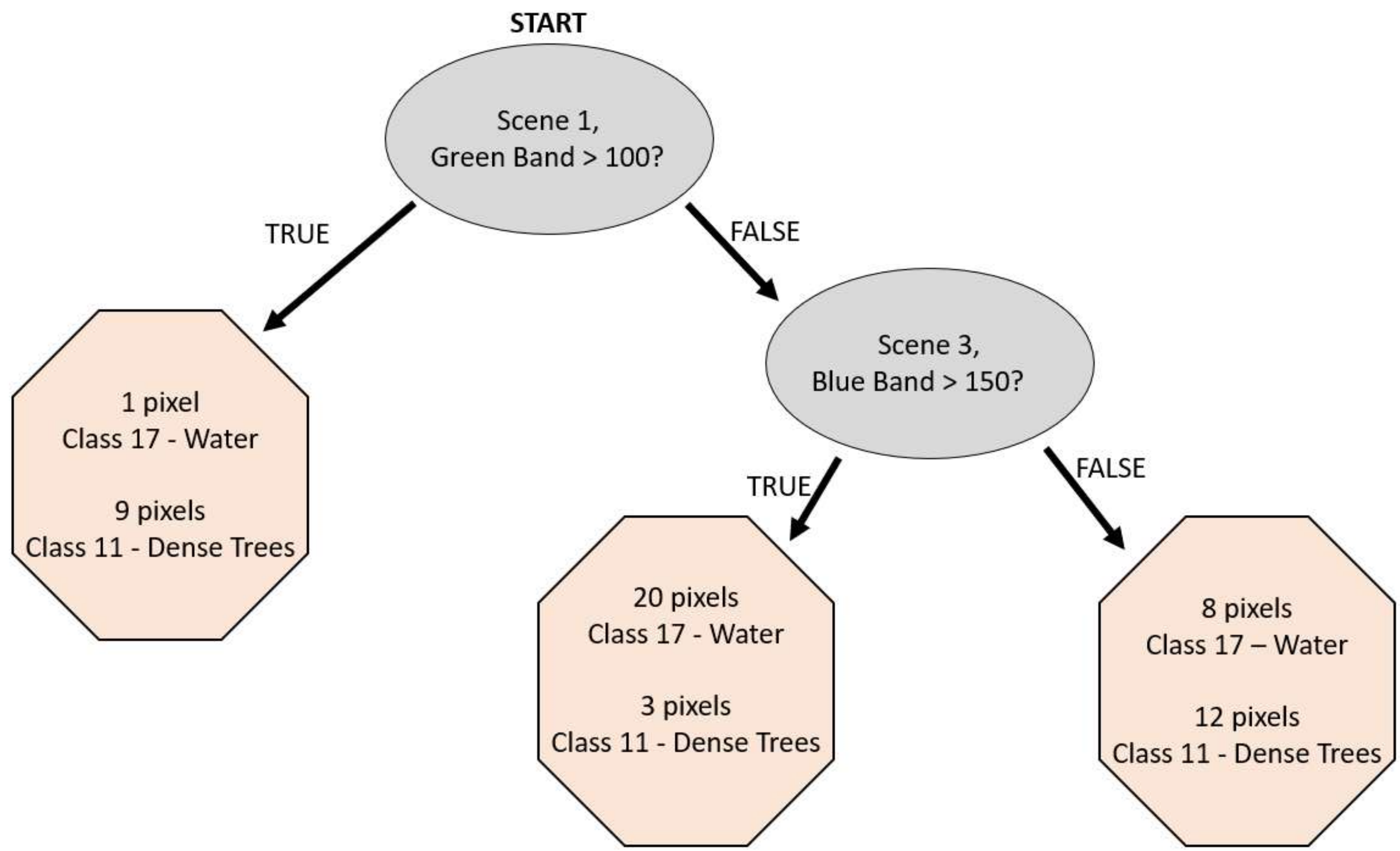


Delineation of training and test data by polygon and pixel.

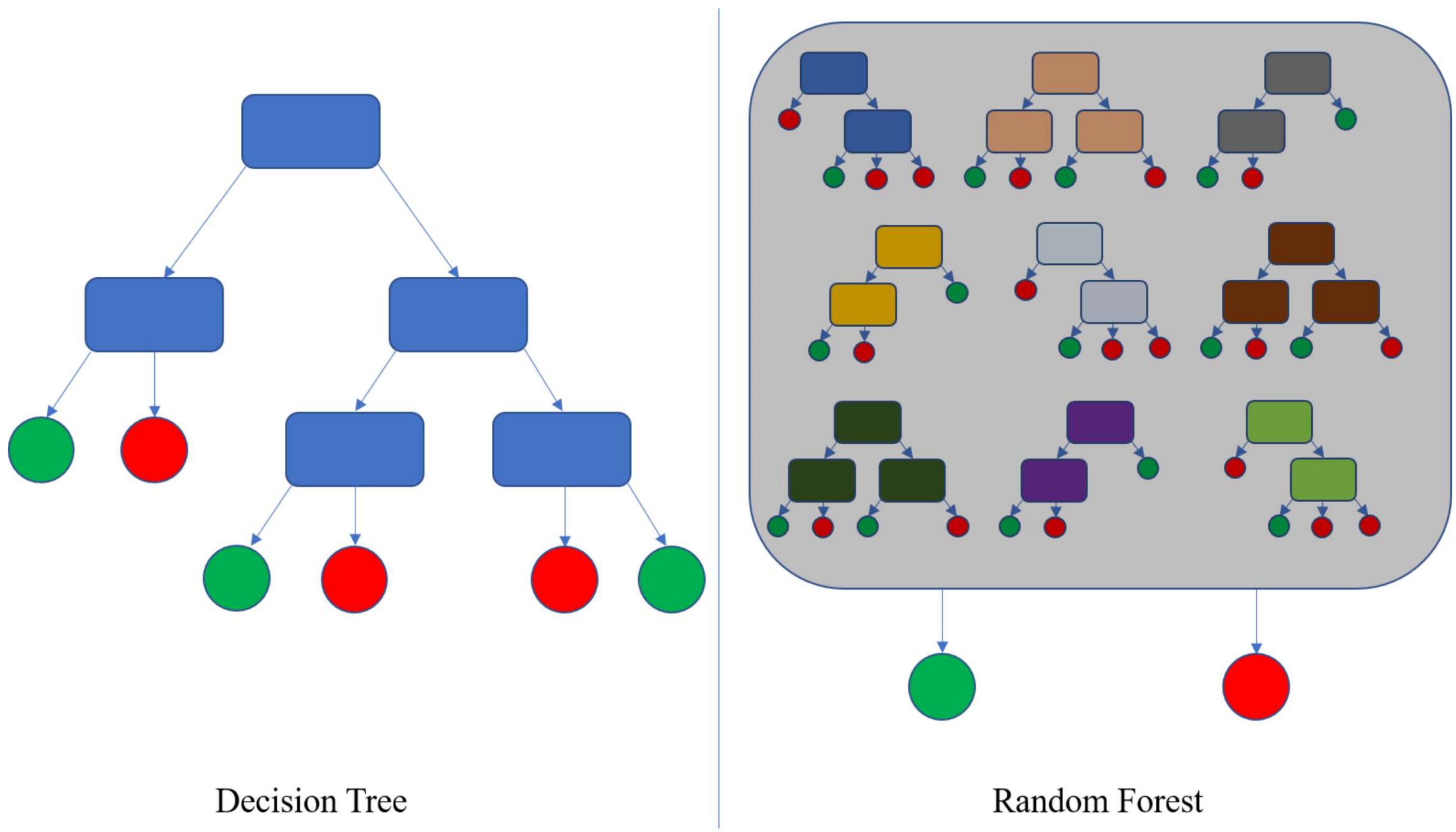
Local Climate Zone	Train	Test
Class 1: Compact high-rise	13 (295)	13 (336)
Class 2: Compact mid-rise	6 (117)	5 (62)
Class 3: Compact low-rise	7 (185)	7 (141)
Class 4: Open high-rise	10 (275)	9 (398)
Class 5: Open mid-rise	4 (79)	4 (47)
Class 6: Open low-rise	6 (60)	7 (60)
Class 7: Lightweight low-rise	0 (0)	0 (0)
Class 8: Large low-rise	4 (90)	5 (47)
Class 9: Sparsely built	0 (0)	0 (0)
Class 10: Heavy Industry	4 (107)	5 (112)
Class 11: Dense trees	7 (762)	7 (854)
Class 12: Scattered trees	6 (194)	7 (213)
Class 13: Bush, scrub	4 (459)	5 (232)
Class 14: Low plants	6 (346)	6 (222)
Class 15: Bare rock or paved	0 (0)	0 (0)
Class 16: Bare soil or sand	0 (0)	0 (0)
Class 17: Water	5 (1266)	5 (1113)

[a] Number of polygons is listed first, with number of pixels in parentheses.

Decision Trees



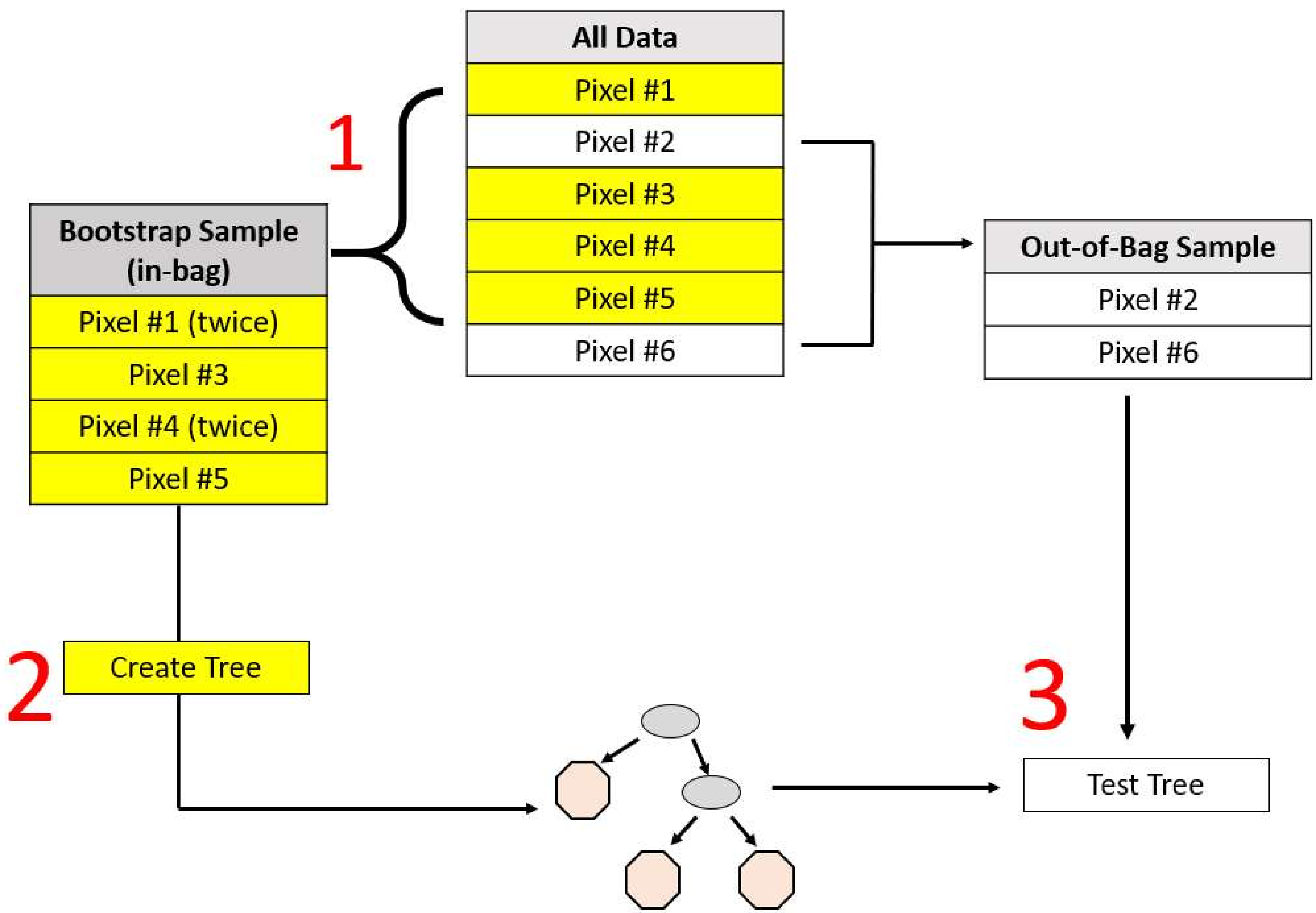
Random Forests: a collection of decision trees



Why is it a [**Random**] Forest?

- Randomizing variables tried at each node
- Bootstrapping samples for each tree

Out-of-Bag Error



Tuning Parameters

$$\text{ntree} = \text{varied}$$

$$\text{mtry} = \sqrt{\#\text{ of parameters}} = \sqrt{36} = 6$$

$$\text{nodesize} = 1$$

$$\text{maxnodes} = \text{maximum possible}$$

Accuracy Assessment

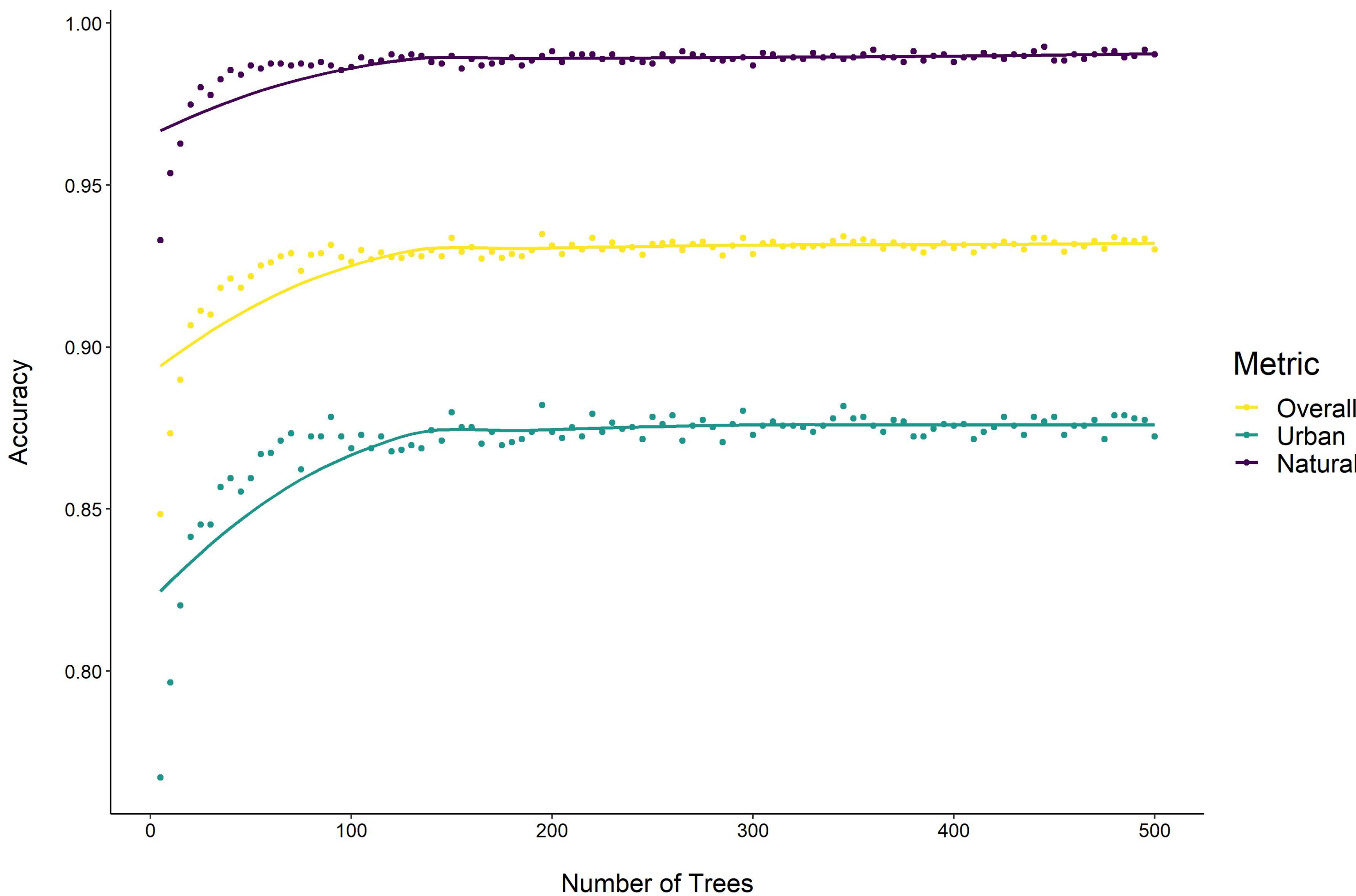
$$\text{Overall Accuracy} = \text{OA} = \frac{\text{number of correctly classified reference sites}}{\text{total number of reference sites}}$$

$$F_1 \text{ Score} = 2 * \frac{UA * PA}{UA + PA}$$

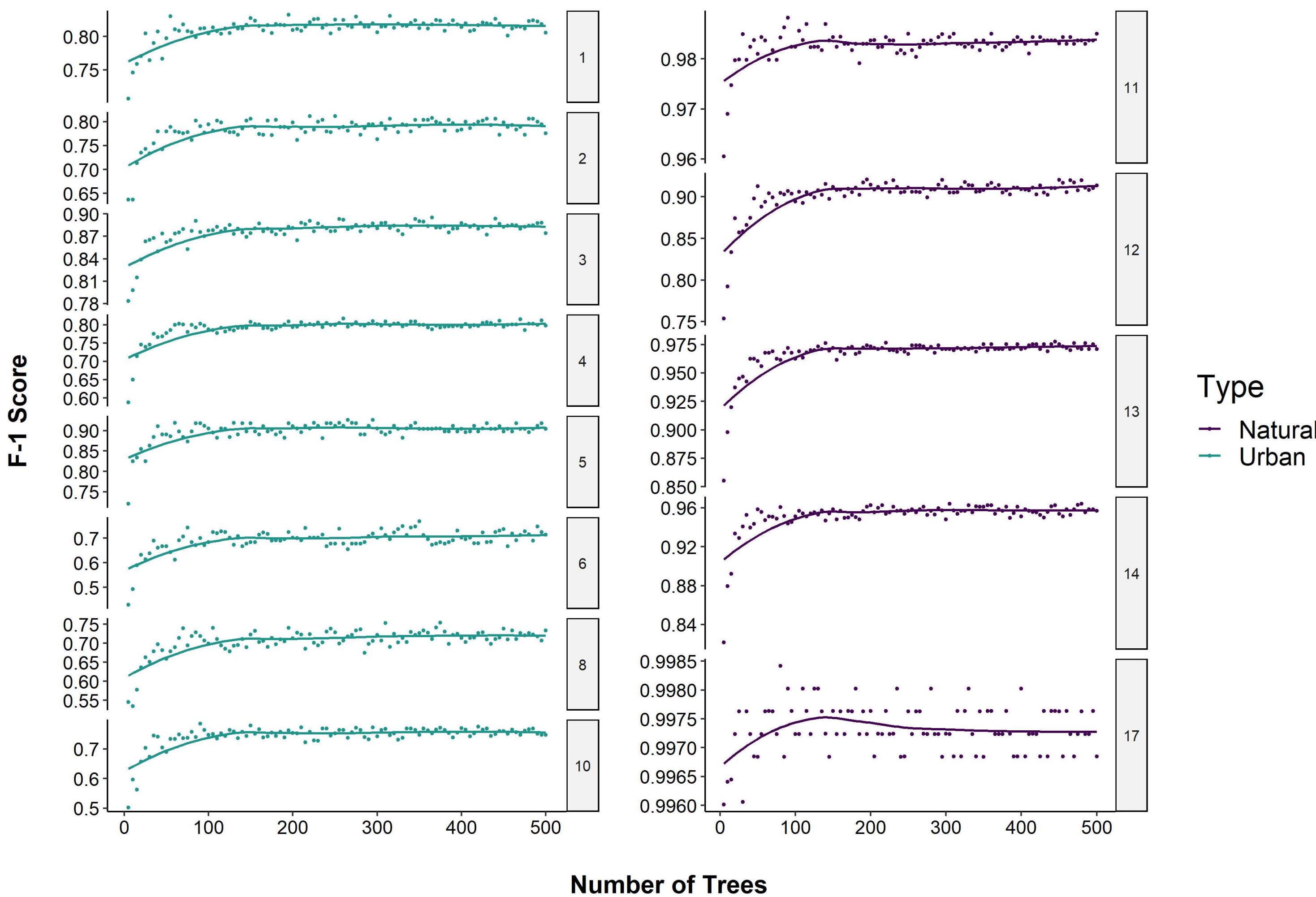
$$UA(z) \ = \frac{\text{number of correctly identified pixels in class z}}{\text{total number of pixels identified as class z}}$$

$$PA(z) = \frac{\text{number of correctly identified pixels in class z}}{\text{number of pixels truly in class z}}$$

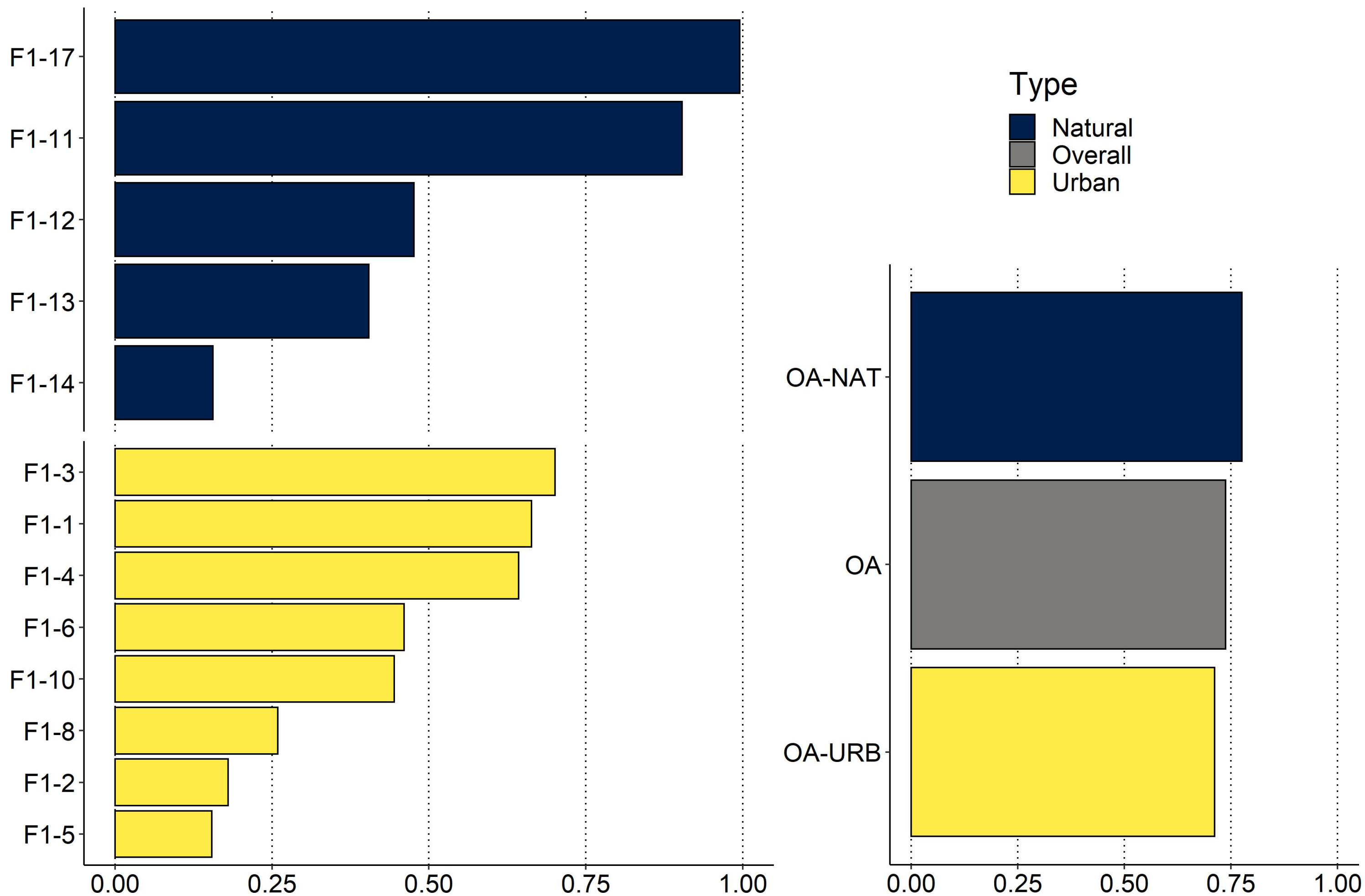
OA Metrics increase as number of trees increase, leveling off around 125 trees. Natural classes have higher OA values.



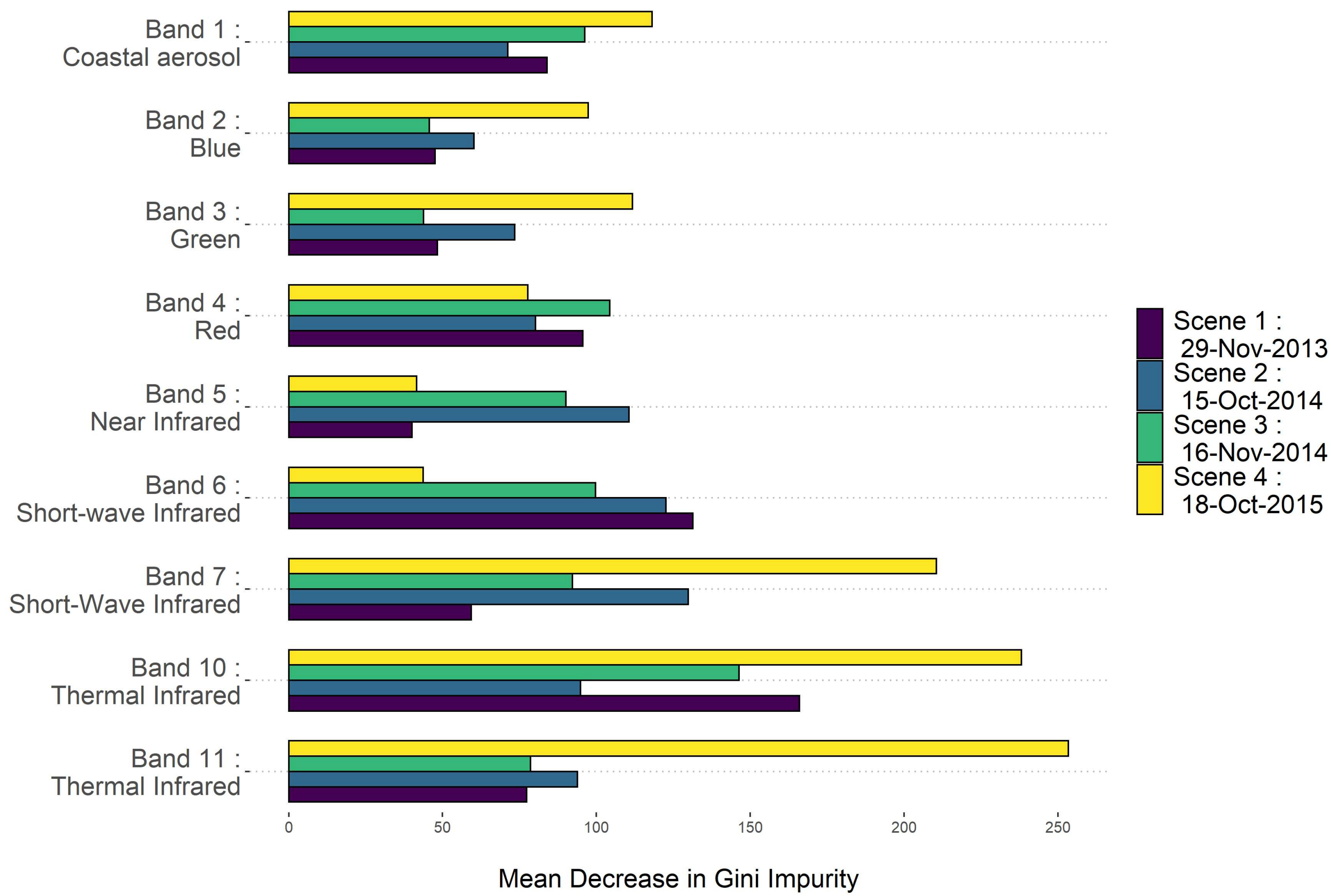
F-1 Score by Class increases as number of trees increases, leveling off around 100 trees.



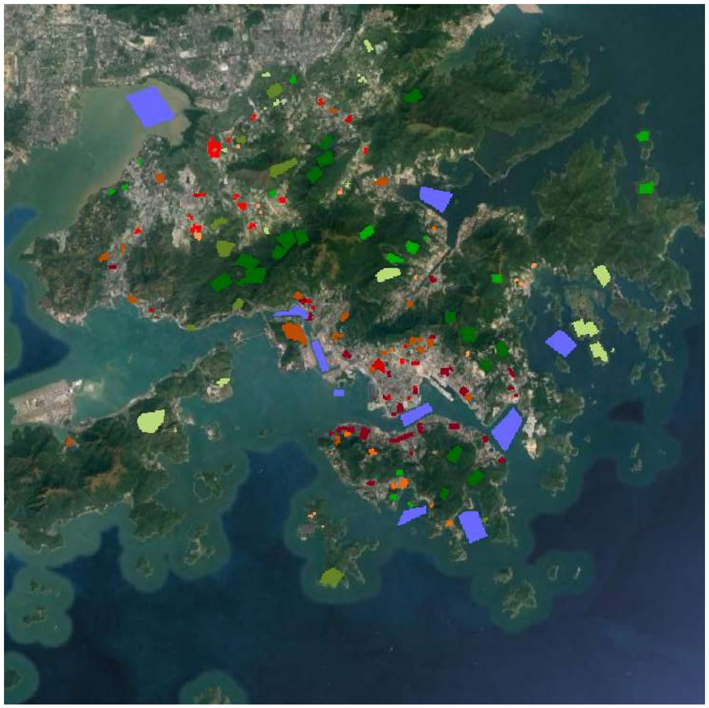
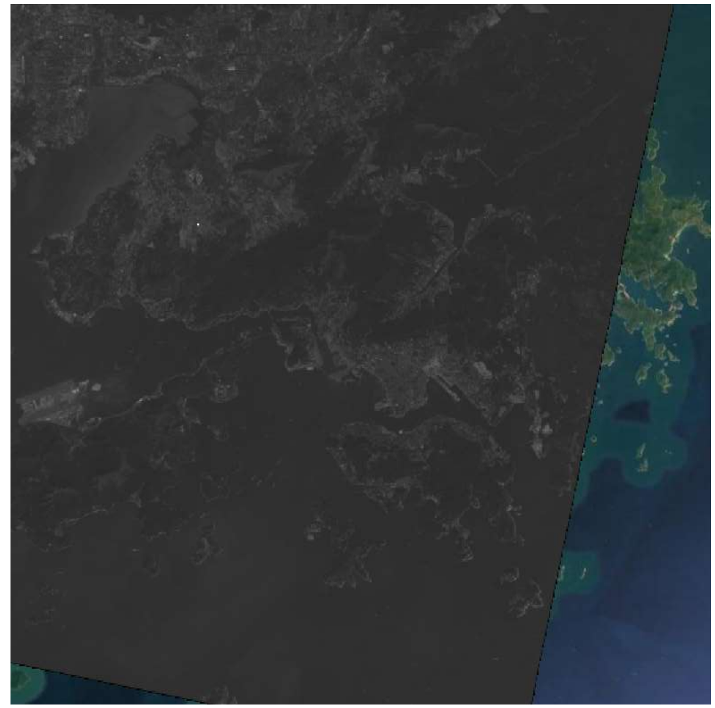
Validation Metrics are much lower for test data than for out-of-bag data.
High OA values may mask low F1 scores.



Importance Measures don't give a clear answer about which predictor variables are most useful.

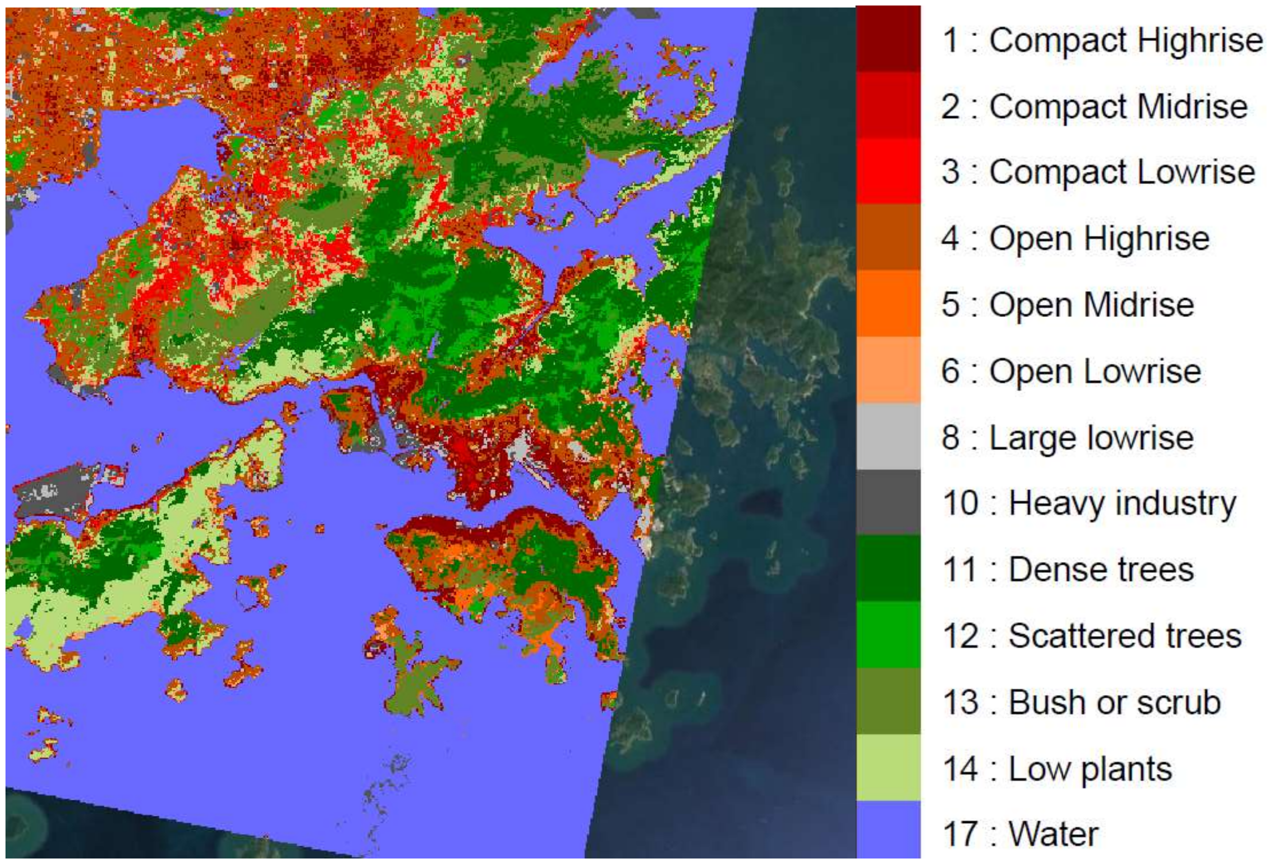


Creating the Full Prediction

LCZ Polygons	Scene 1, Band 4

Creating the Full Prediction



Conclusion

Overall Results:

- Low accuracy for prediction on the test data, in comparison to the out-of-bag data
- High OA values can mask low F1 scores within classes

Limitations:

- Reference polygons on account for ~3% of the Area of Interest
- Time constraints

Future Work:

- Multiple tuning parameters & the interactions between them
- Quantifying "how much" ground truth data is enough

Acknowledgements

Questions?

All code and higher resolution images for this project can be found on GitHub at https://github.com/erickabsmith/masters-project-lcz-classification.