The Impacts of Training Data Spatial Resolution on Deep Learning

LTC Christopher Ardohain

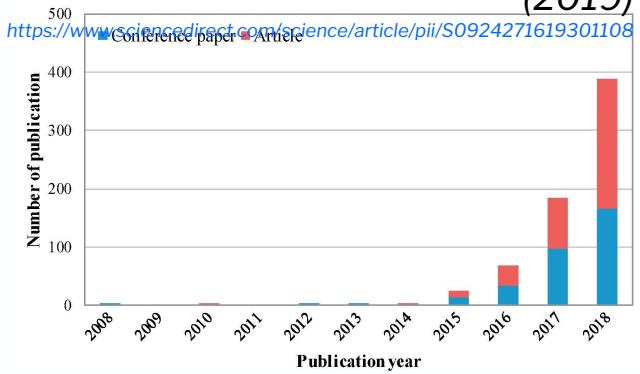
11/07/2024

PURDUE UNIVERSITY
Forestry and Natural Resources

The Problem

- Deep Learning is becoming ubiquitous, especially for remote sensing analysis
- Deep Learning is data hungry, but the development of suitable training datasets is time consuming and expensive
- These costs require us to explore alternative methods for training data development
- A potential method is the application of training data across spatial scales, but the impact of such application has yet to be quantified
- We compare 3DEP(1.5m) to NLCD(30m) derived forest boundary training data to quantify the impact on deep learning model performance

From *Deep learning in remote sensing applications: A meta-analysis and review (2019)*

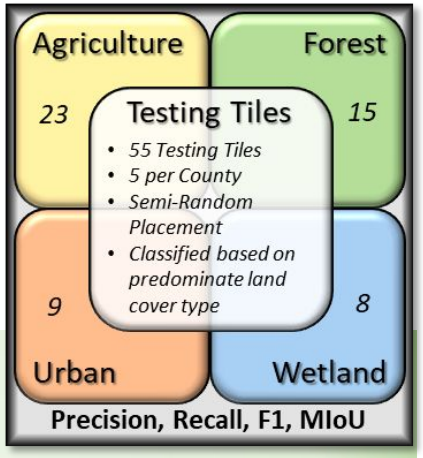


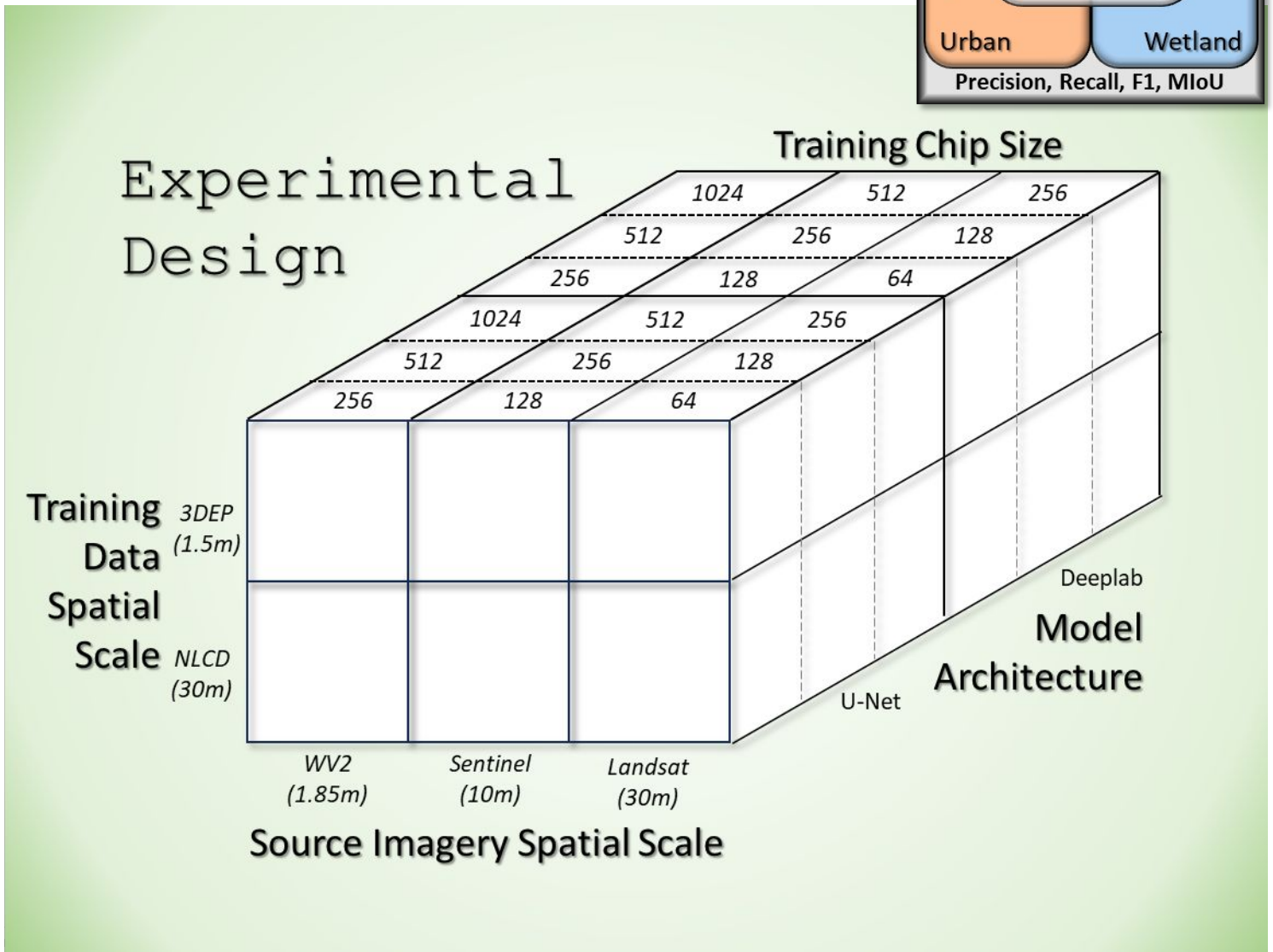
3DEP Generated **NLCD**

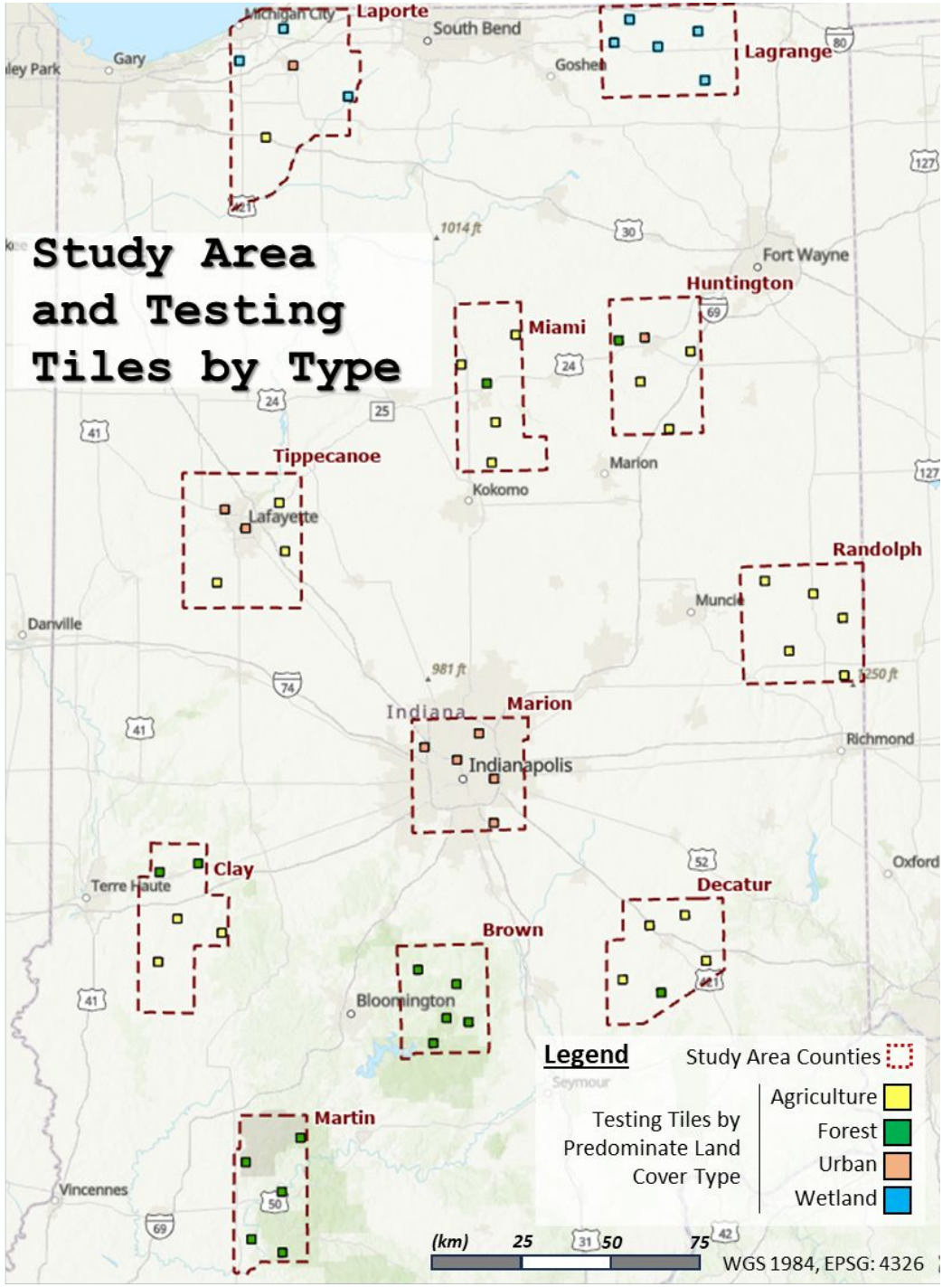
Training Datasets

PURDUE UNIVERSITY
Forestry and Natural Resources

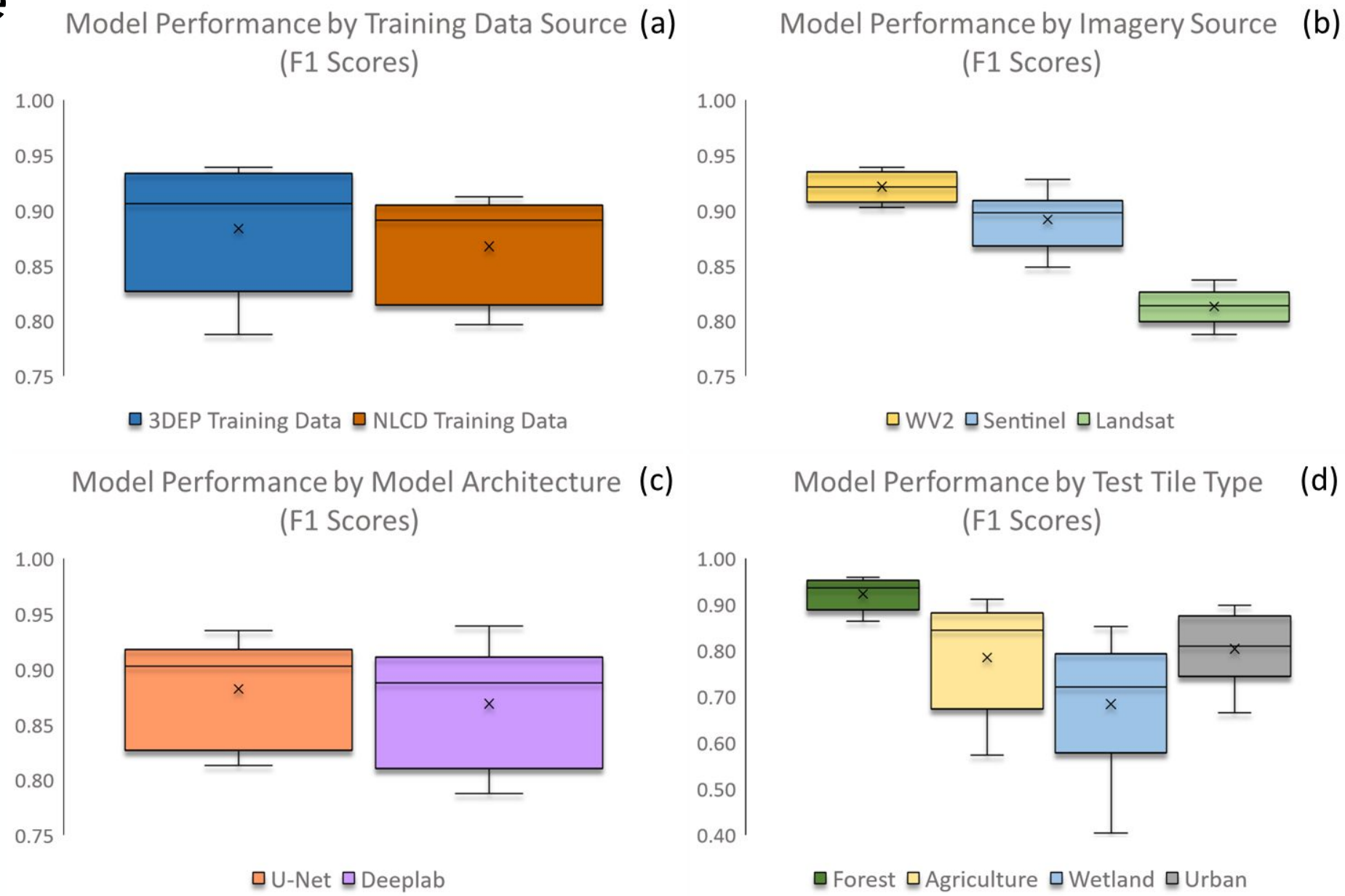
Data and Methods



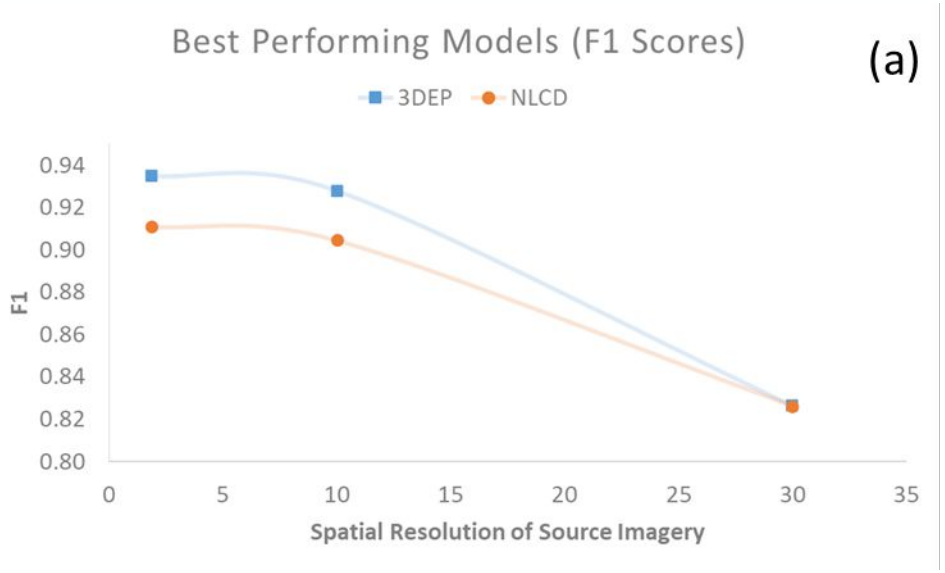


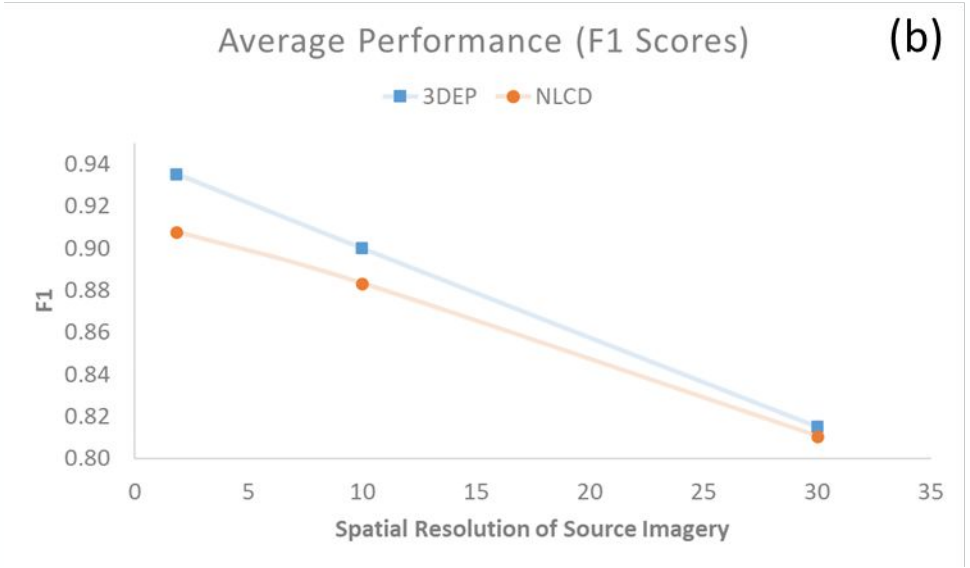


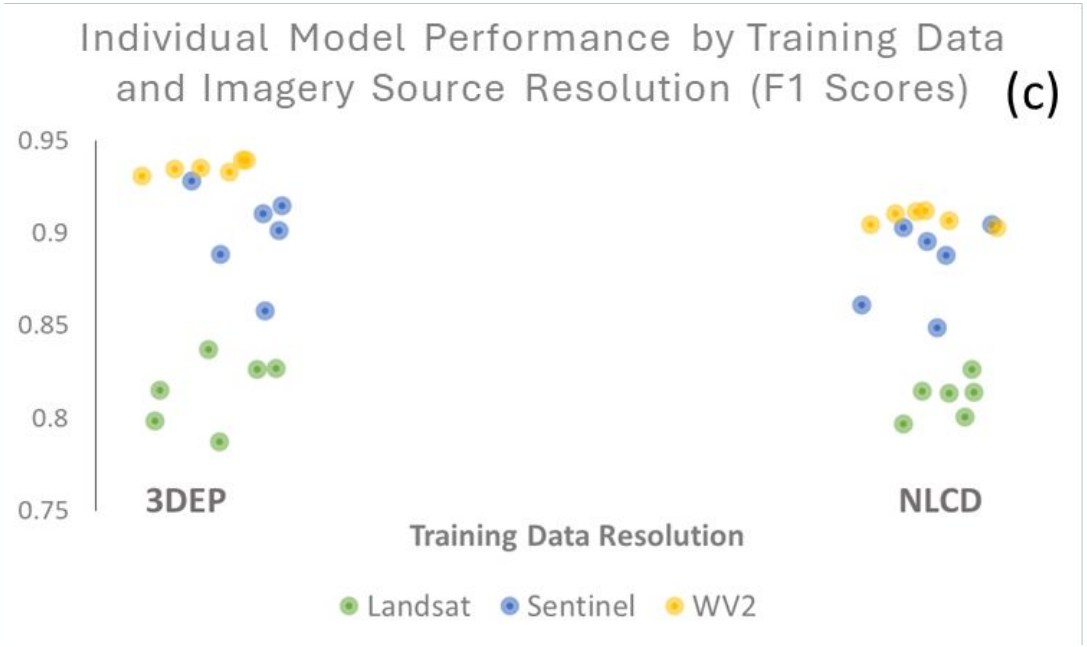
Discernable Patterns in Model Performance – Univariate Analysis



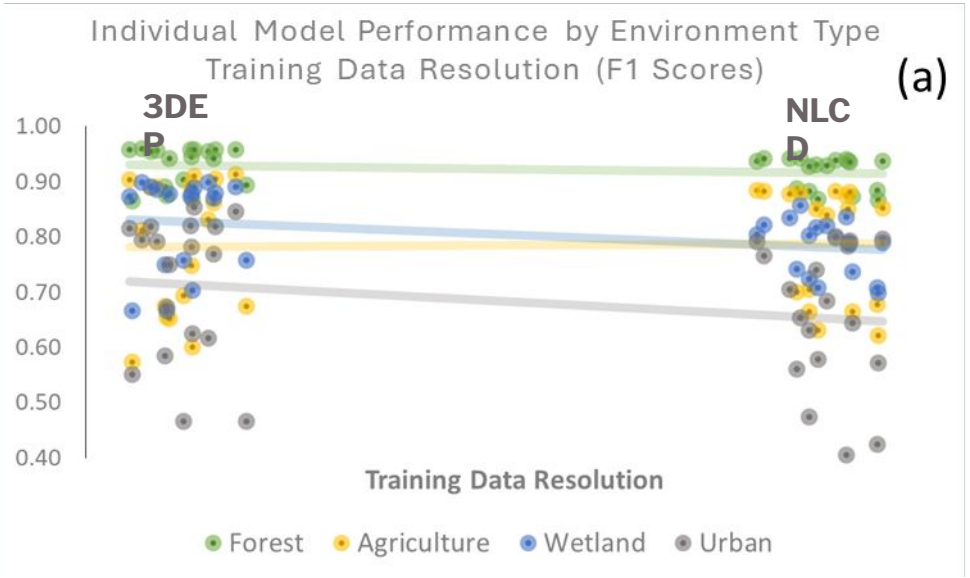
Relationship between Training Data and Source Imagery Spa

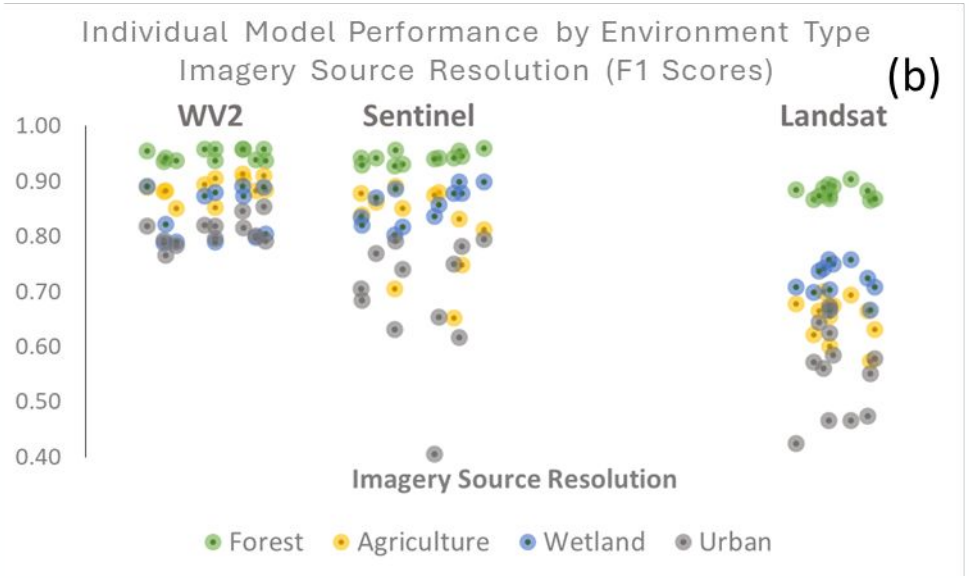


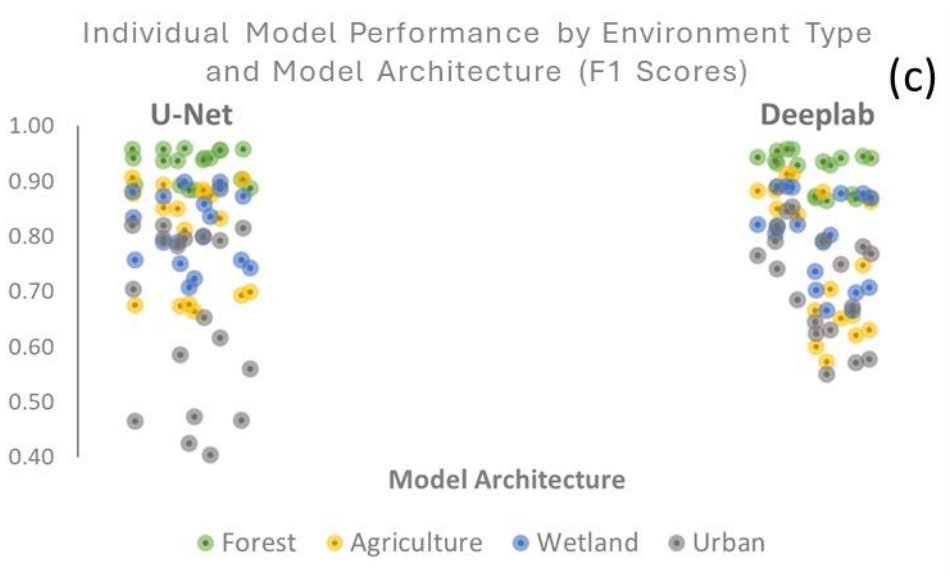




Impact of Environment Type







Best Model Performance (F1) per Tile Type

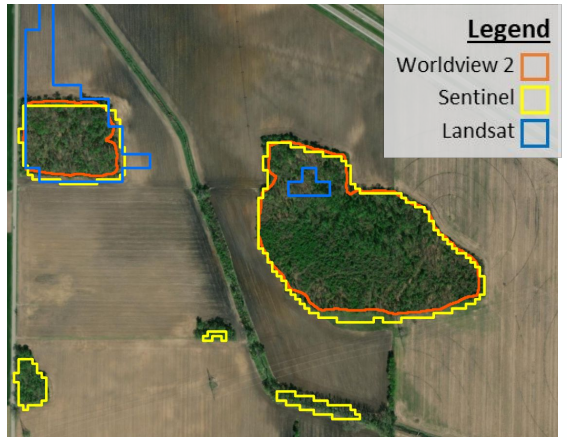
Agriculture Tiles

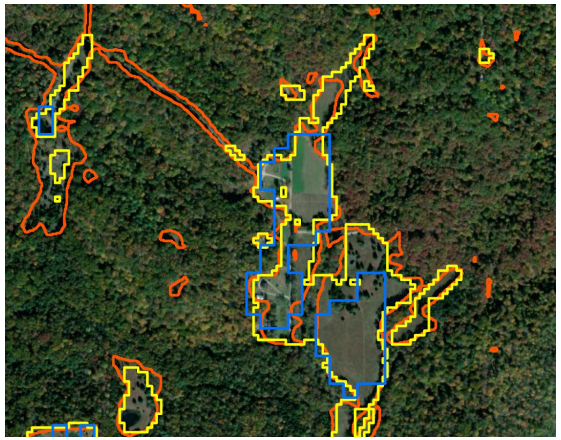
Forest Tiles

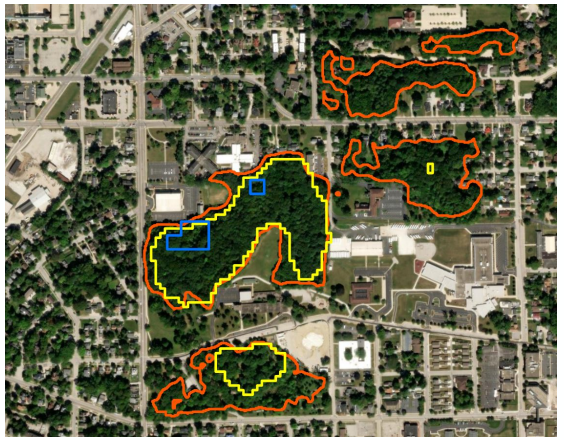
Urban Tiles

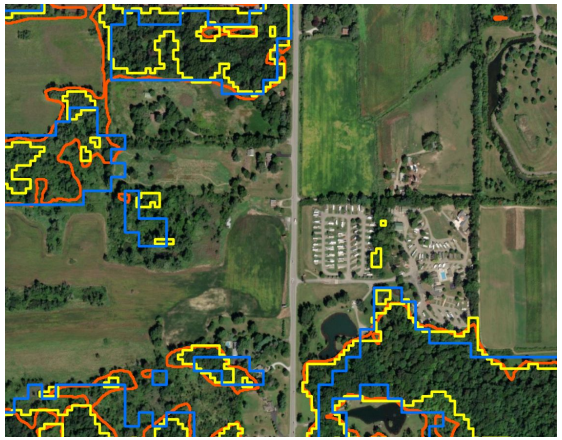
Wetland Tiles

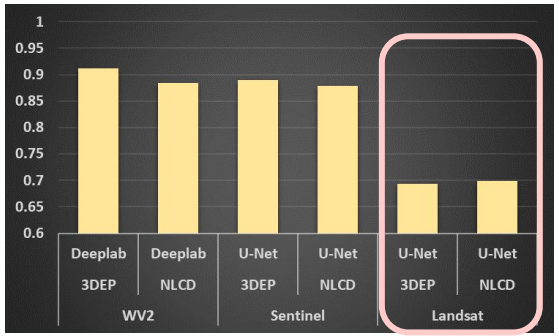
3DEP Training Data

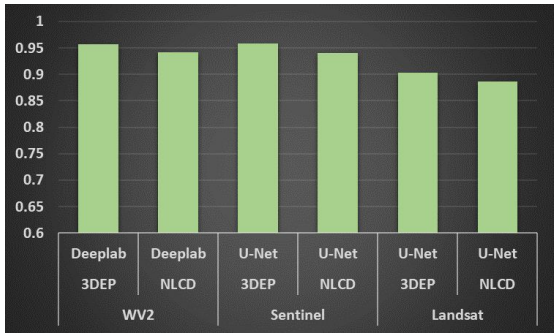


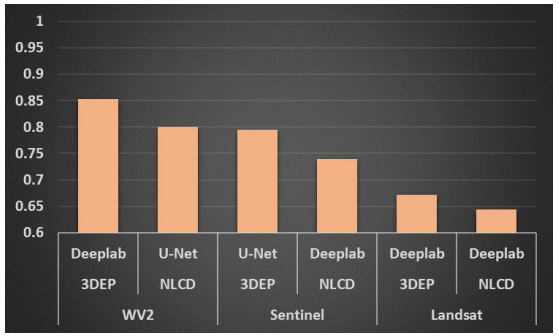


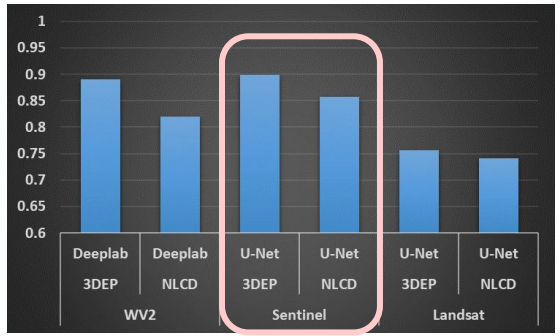




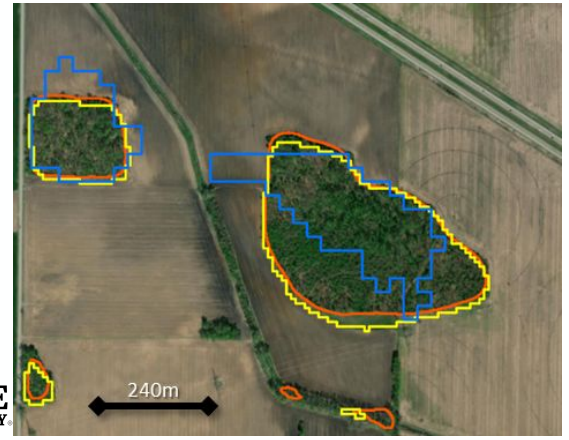


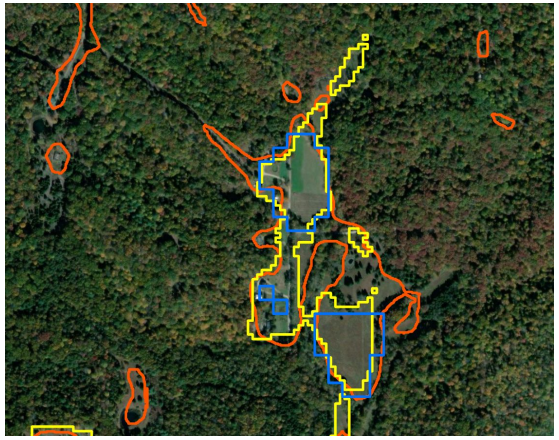


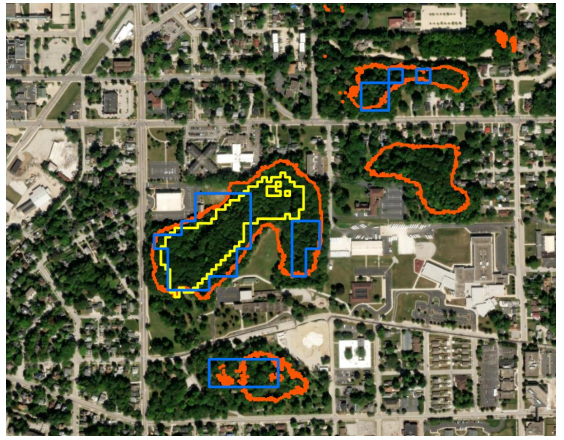


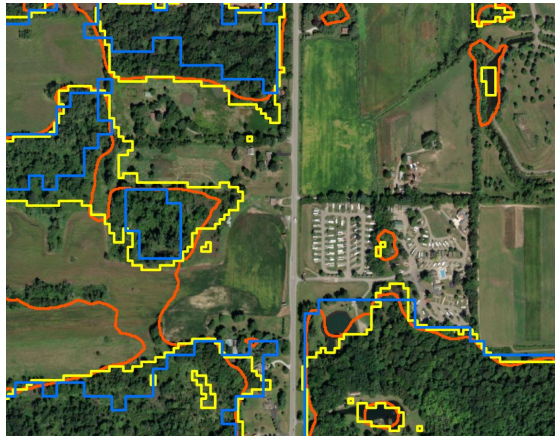


NLCD Training Data









PURDUE UNIVERSITY
Forestry and Natural Resources

So What…?

- Higher spatial resolution training data produced more accurate models regardless of imagery source spatial resolution, however, the gap in model performance (F1) was only ~2.7% even at its most extreme.

- Performance based on land cover varied greatly from average F1 scores of 0.923 in homogenous forested areas to 0.684 in complex urban environments

- Although the results show no difference in training time between data sources, training data chipping with 3DEP annotations took roughly 5 times longer.

- Other observations
  - Training Chip Size: Sentinel source imagery was the only data subset strongly impacted by training chip size (smaller training chips produced better results)
  - Deeplab was much more efficient at training than U-Net but performed slightly worse
  - Model accuracy relationships remained intact when total number of training chips was held constant for all imagery sources (3 additional U-Net models tested at a chip size of 256 w/ 3848 total chips)