Deriving Vegetation Variables from Satellite Observations using a Data-driven Approach

Alan Wang
Mentor: Heshun Wang
Cooperative Institute for Satellite Earth System Studies

Objectives

Background:
- fCover measures the spatial extent of vegetation.
- fCover is critical to understanding the climate-vegetation feedback and land-use change.

Problem: Ground-based techniques for measuring fCover are limited in scope, while satellite-based techniques provide global coverage but need calibration/validation and quality control.

Purpose: Apply ML frameworks/models to generate predictions of ground-measured fCover based on globally-available VIIRS data.

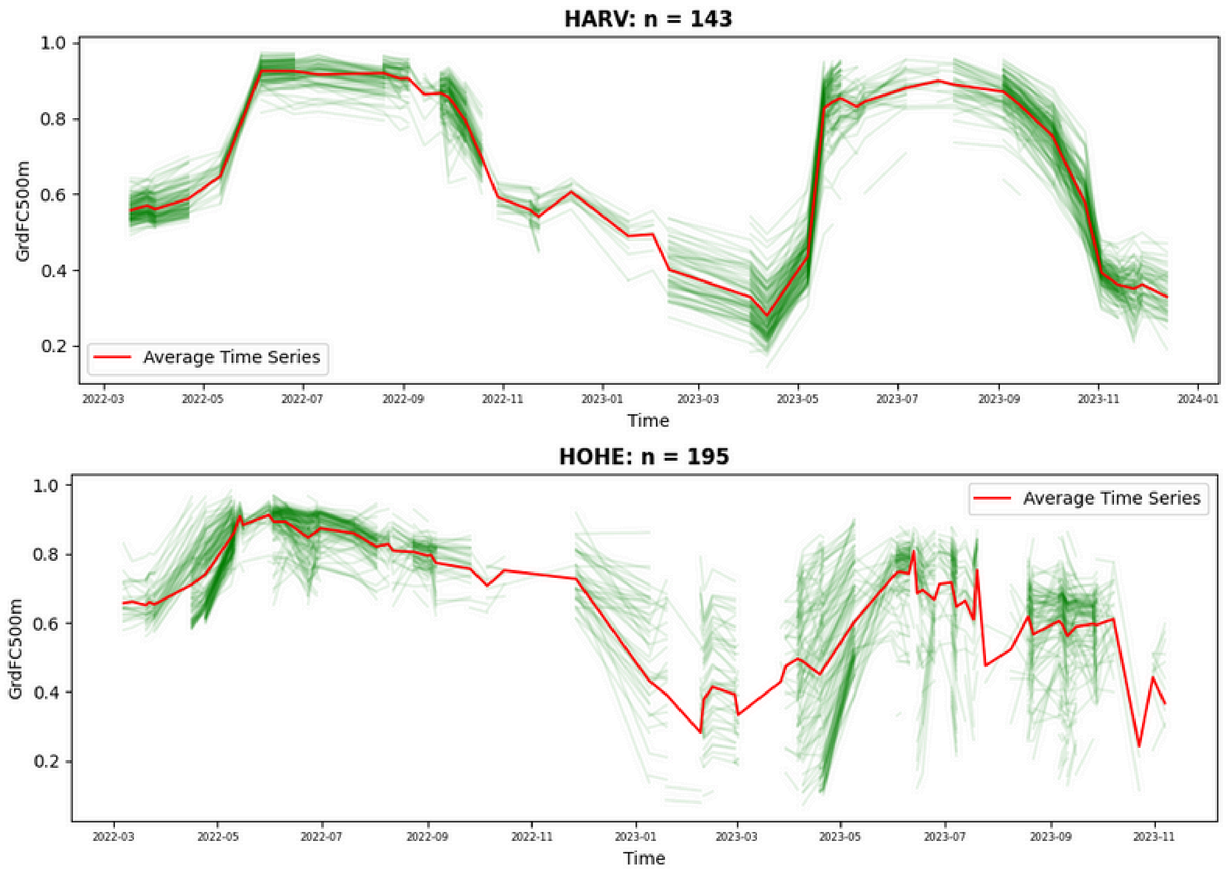

Fig 1: fCover GBOV data over two ground measurement sites

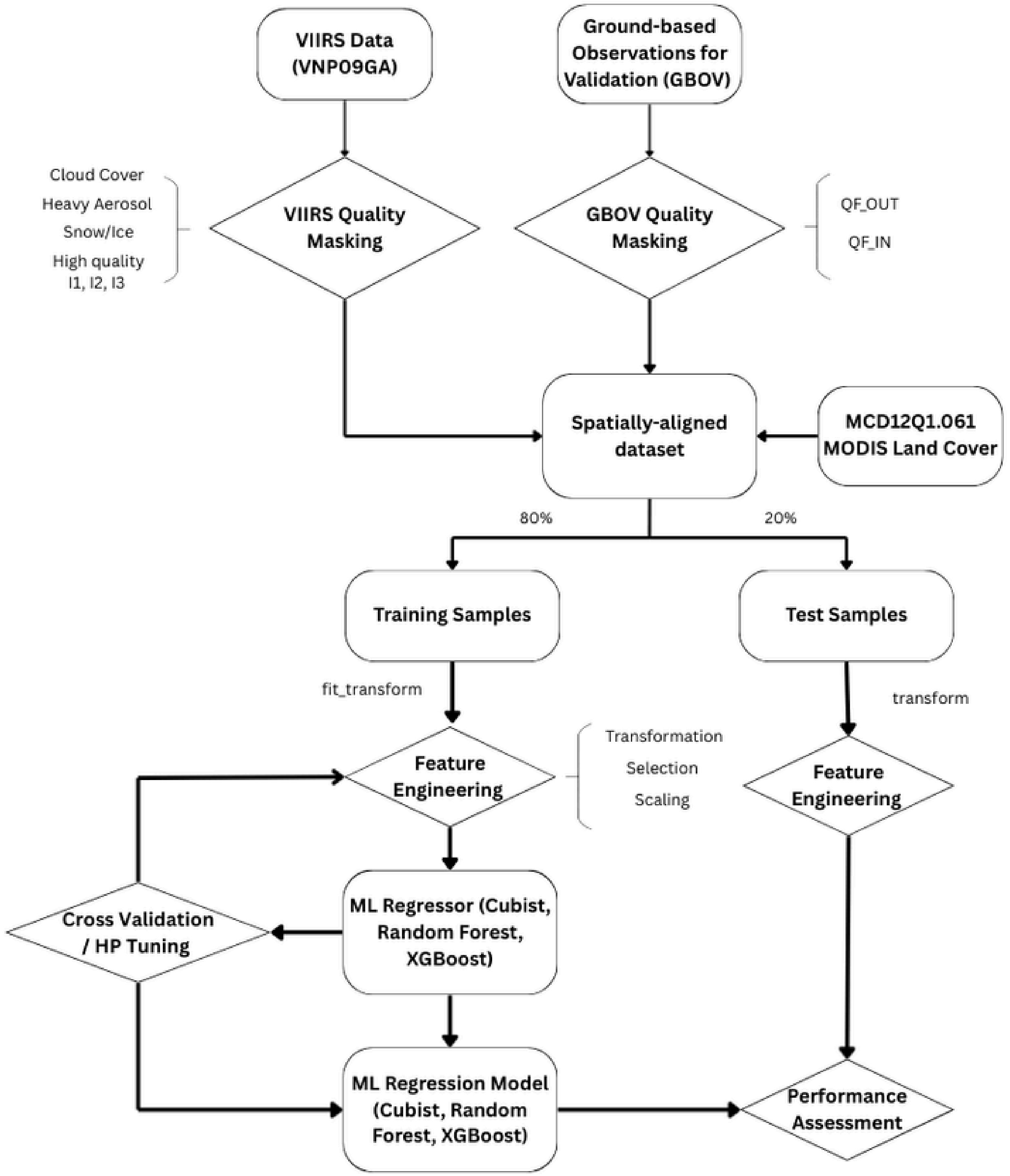

Fig 3: Illustration of development process for ML models.

Methodology/Data

Datasets:
- VNP09GA: VIIRS Surface Reflectance
- MCD12Q1.061 MODIS
- GBOV

Procedure:
- Spatially align VIIRS feature information, landcover, and GBOV.
- Perform data preprocessing and feature engineering/selection.
- Assess performance of XGBoost, random forest, and cubist regression models using K-fold stratified cross sampling with MAE and R-Squared as metrics.

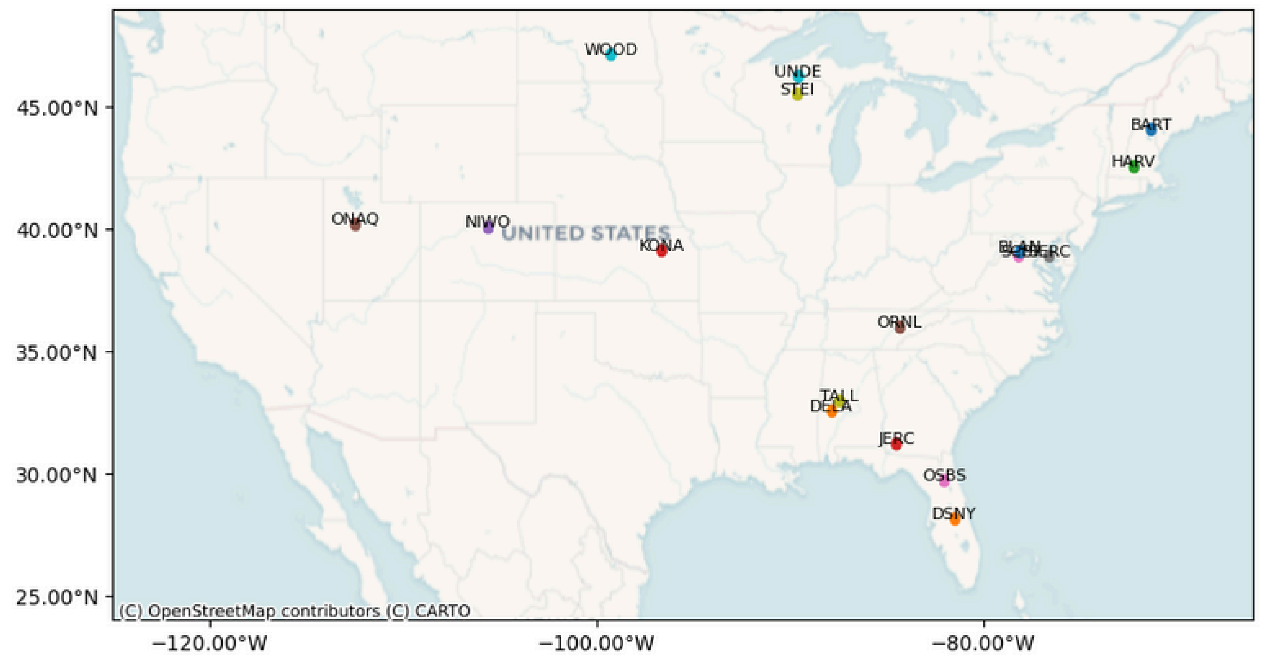

Fig 2: NEON Sites for fCover GBOV across continental US

Deriving Vegetation Variables from Satellite Observations using a Data-driven Approach

Alan Wang

Mentor: Heshun Wang

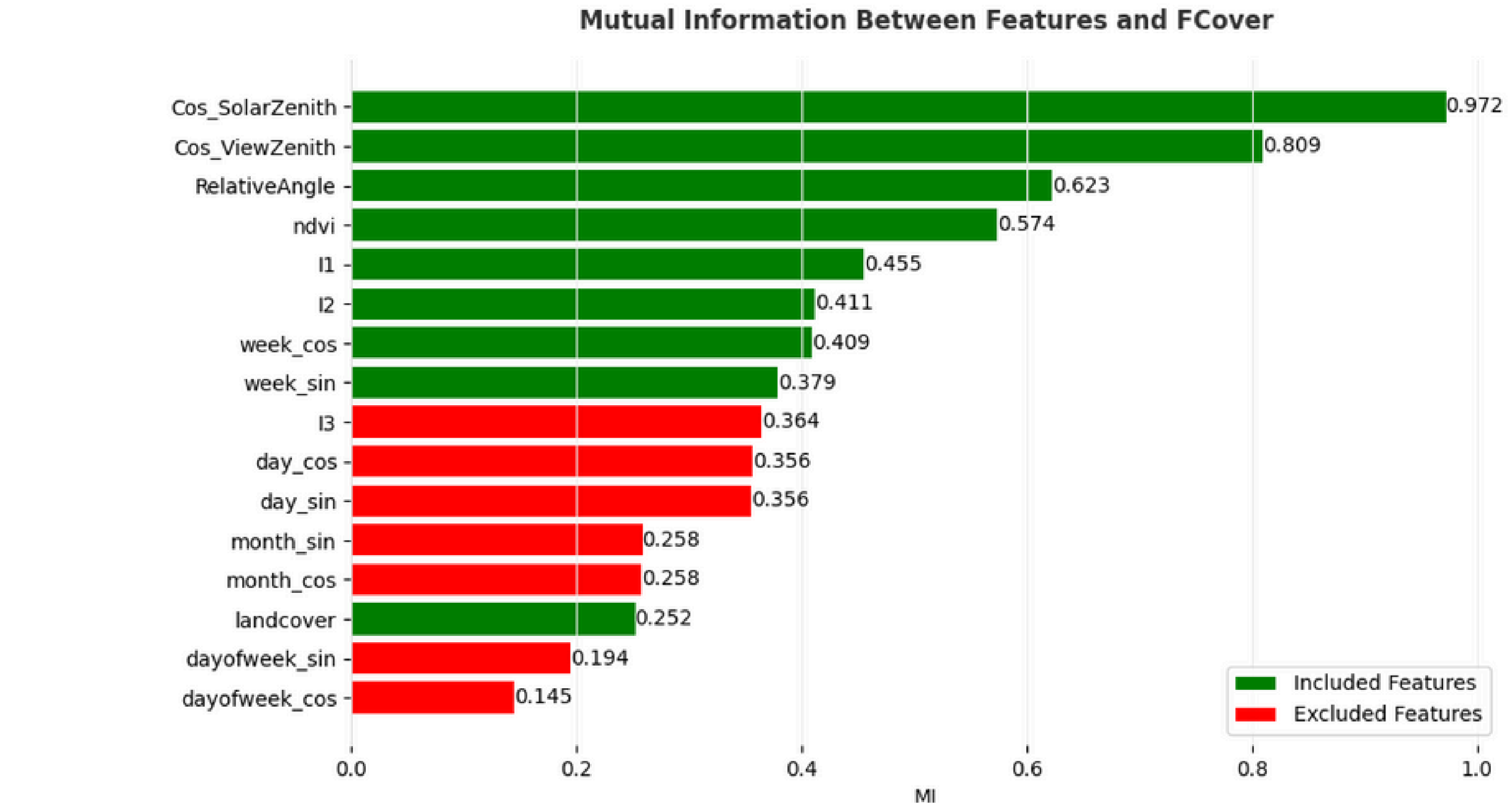

Fig 4: Feature selection based on mutual infomration and MDA

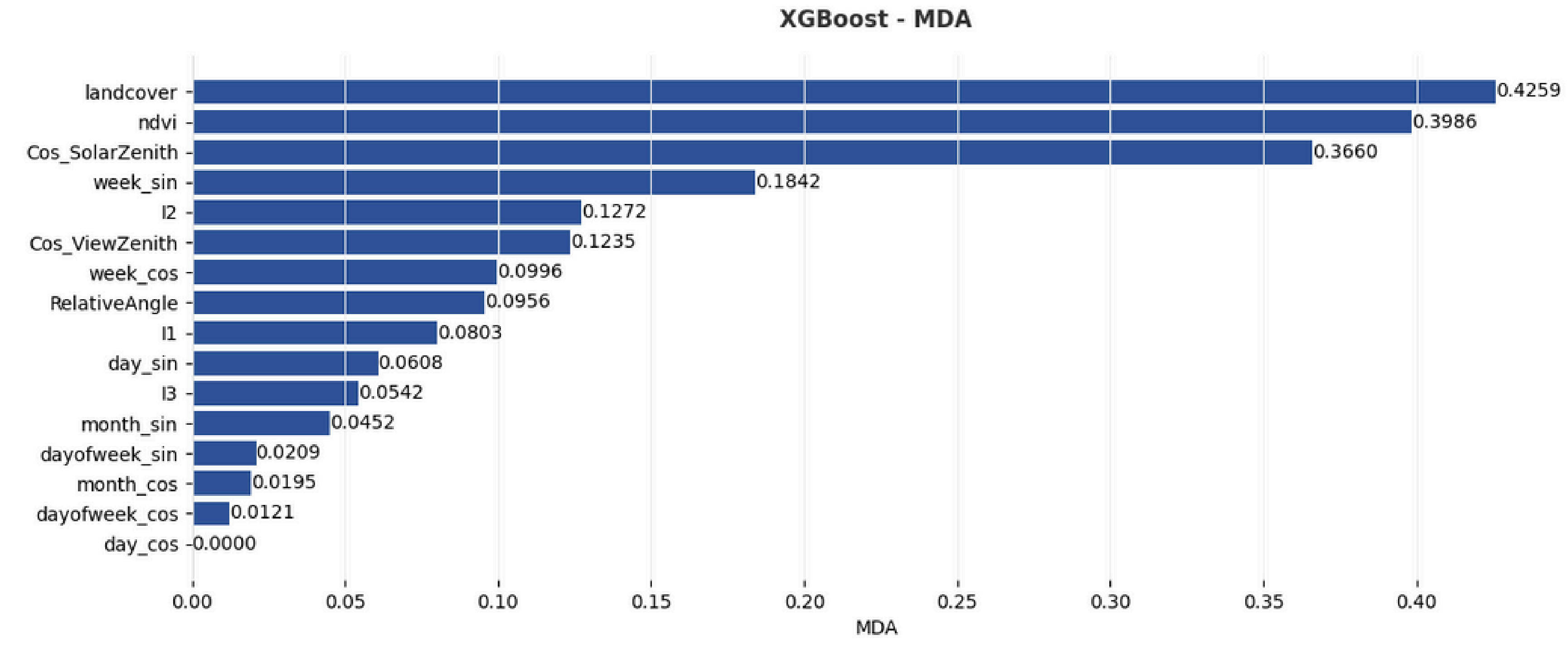

Fig 5: XGBoost permutation feature importance (MDA)

Table 1: Model performances across stratified k-fold cross validation

	Fold 1 (R^2)	Fold 2 (R^2)	Fold 3 (R^2)	Fold 4 (R^2)	Fold 5 (R^2)	Mean (R^2)	SD (R^2)
Cubist	0.888633	0.889113	0.890394	0.891903	0.891903	0.890909	0.0013799
XGBoost	0.892615	0.887513	0.89036	0.892356	0.893786	0.891326	0.002202
Random Forest	0.895547	0.894318	0.894804	0.897094	0.900401	0.896433	0.002195

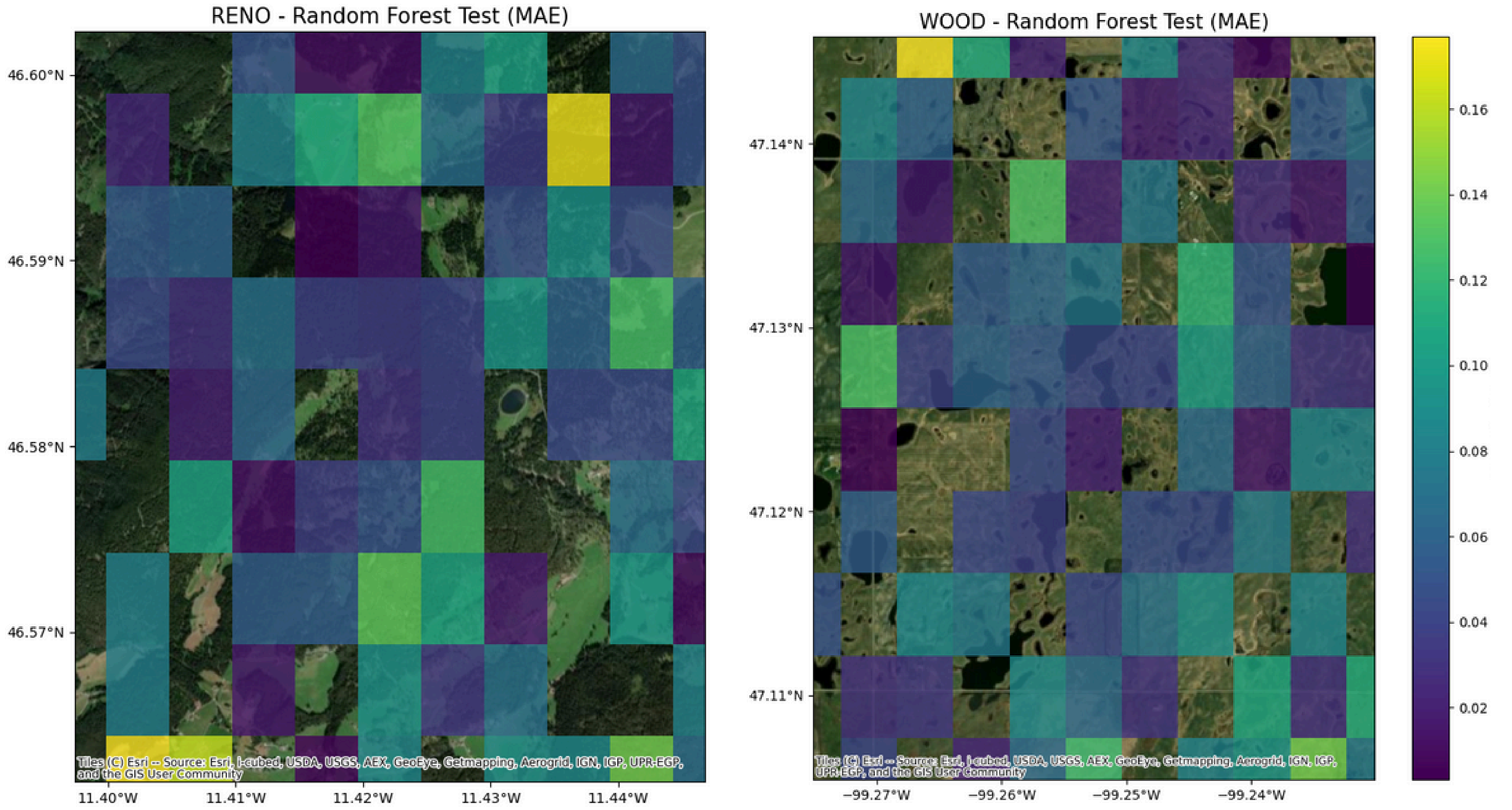

Fig 6: Visualization of random forest performance on test set over two ground measurement sites

Deriving Vegetation Variables from Satellite Observations using a Data-driven Approach

LSTM

Alan Wang

Mentor: Heshun Wang

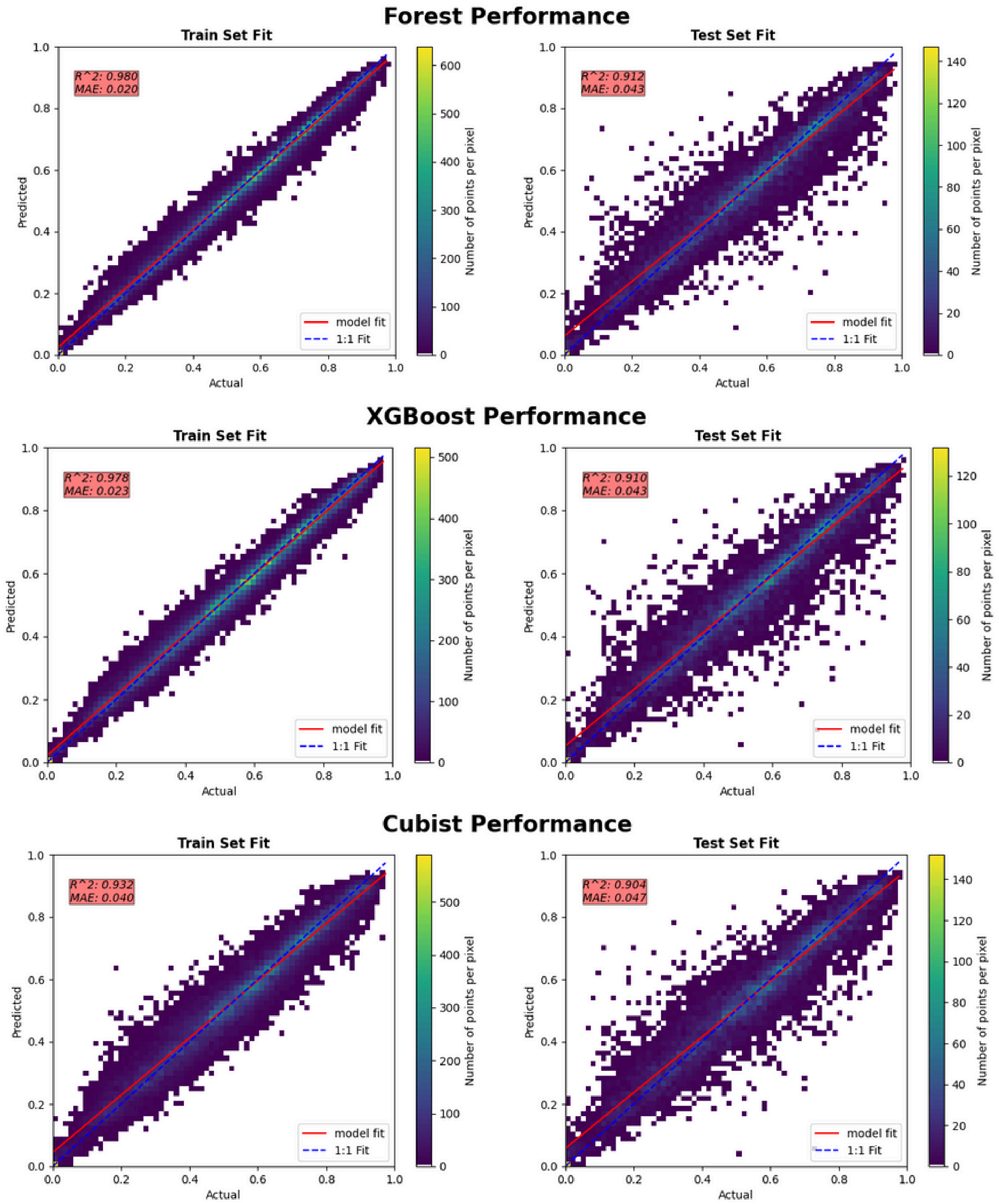

Fig 7: Model performances on training set vs test set

Results

- Random forest model outperformed the cubist and random forest regressor models over the test set and the K-fold cross validation, however all three models had comparable performances.
- Performance of models were consistent over different folds of the training set.
- All three models were able to capture the annual growth cycle present in the fCover GBOV.
- All three models highly valued landcover, NDVI (derived from I1/I2 bands), and angle information (Cos_SolarZenith and RelativeAngle)

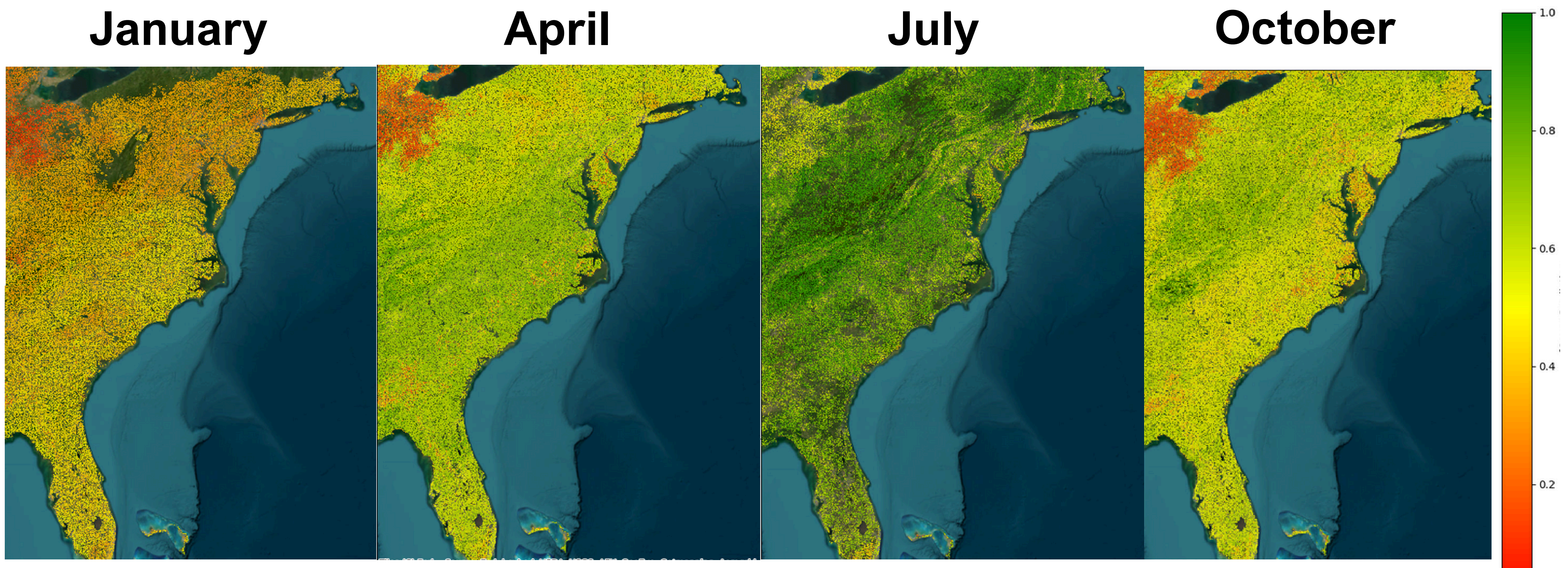

Fig 8: Random forest regressor model median predictions over East Coast for 2023