# **Greenness Prediction**

- Target Variables (#1): gcc 90 Features Used (#19): gcc90 value from (t-7)th to (t+7)th day of last year and weather variables of the current day (max\_temp, min\_temp, radiation, precipitation)
- Model used: Random Forest Regressor
- Cross-validation strategy used: Time Series Cross Validator (n splits = 3)
- Train Data: Till 2019
- Test Data: 2020
- Note
  - Model trained separately for each of the sites
  - Hyperparemeters of Random Forest Regressor are also different for each of the sites
  - The site UKFS does not have sufficient data. Based on the assumption that other sites are working well with this model, UKFS model has been trained at the last separately with 2019 and 2020 data without testing.

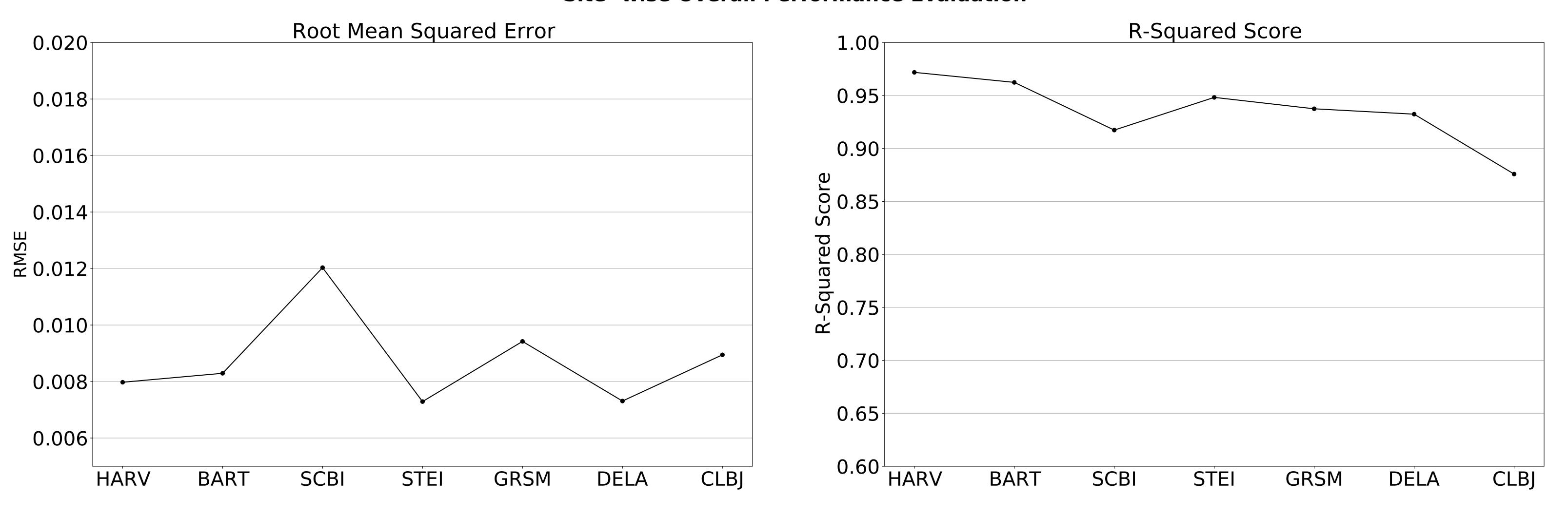
# **Greenness Prediction**

- Target Variables (#1): gcc 90
- Features Used (#19): gcc90 value from (t-7)th to (t+7)th day of last year and weather variables of the current day (max\_temp, min\_temp, radiation, precipitation)
- Model used: Random Forest Regressor
- Cross-validation strategy used: Time Series Cross Validator (n splits = 3)
- Train Data: Till 2019
- Test Data: 2020
- Note
  - Model trained separately for each of the sites
  - Hyperparemeters of Random Forest Regressor are also different for each of the sites
  - The site UKFS does not have sufficient data. Based on the assumption that other sites are working well with this model, UKFS model has been trained at the last separately with 2019 and 2020 data without testing.

### Year-wise no. of observations (after creating features and dropping missing values)

	HARV	BART	SCBI	STEI	UKFS	GRSM	DELA	CLBJ
2016	nan							
2017	10.0	11.0	12.0	nan	nan	nan	9.0	nan
2018	303.0	272.0	296.0	36.0	nan	242.0	311.0	18.0
2019	346.0	299.0	253.0	340.0	41.0	350.0	361.0	327.0
2020	272.0	271.0	331.0	334.0	253.0	298.0	325.0	357.0
2021	nan							

### **Site- wise Overall Performance Evaluation**



### **Feature Importances**

