Using Machine Learning to Detect Cloud Signatures in COSMIC-2 Radio Occultations

Thomas Connor (AER), Jeana Mascio (AER), Stephen S. Leroy (AER), Robert P. d'Entremont (AER), and E. Robert Kursinski (Planet iQ)

Objective

Explore the feasibility of using Machine Learning (ML) to detect clouds using Radio Occultation atmospheric profile data.

Approach

The data used in this study comes from three primary sources:

Inputs

1. Cosmic2 atmprf files
   1. Refractivity Profile
   2. Dry Temperature Profile
   3. Bending Angle Profile
   4. Altitude Profile
2. GFS NWP data
   1. Temperature Profile
   2. RH
3. GOES
   1. Use GOES channel data and NWP to map clouds to voxel locations
   2. Voxels are GOES pixels by 500m deep vertical sections from 6 to 12 km

Target for ML is cloud vs no cloud for voxels that the RO transect intersects

Average Refractivity vs. Layer Middle Height (m)

RH vs. Layer Middle Height (m)
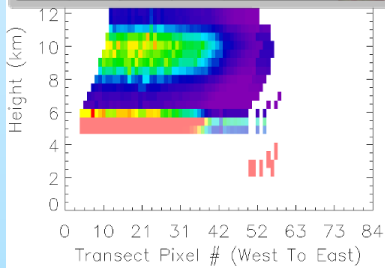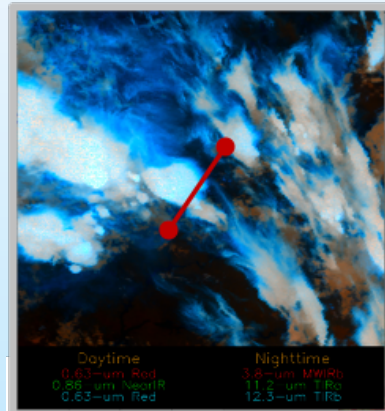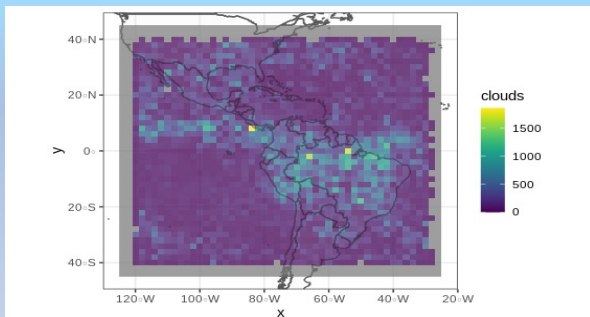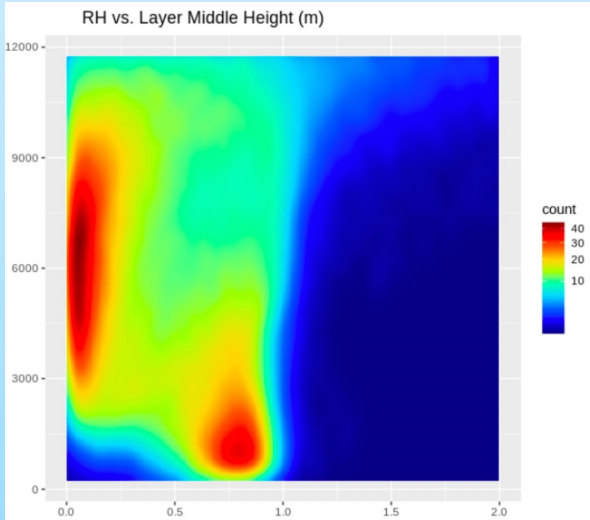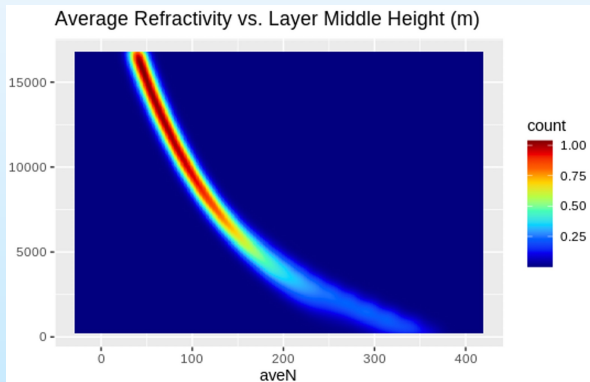

```
Confusion Matrix and Statistics

          Reference
Prediction     0     1
         0 22738  2562
         1  2414  4776

               Accuracy : 0.8468
                 95% CI : (0.8429, 0.8507)
    No Information Rate : 0.7741
    P-Value [Acc > NIR] : < 2e-16

                  Kappa : 0.5589

 Mcnemar's Test P-Value : 0.03717

            Sensitivity : 0.9040
            Specificity : 0.6509
         Pos Pred Value : 0.8987
         Neg Pred Value : 0.6643
             Prevalence : 0.7741
         Detection Rate : 0.6998
   Detection Prevalence : 0.7787
      Balanced Accuracy : 0.7774

       'Positive' Class : 0
```
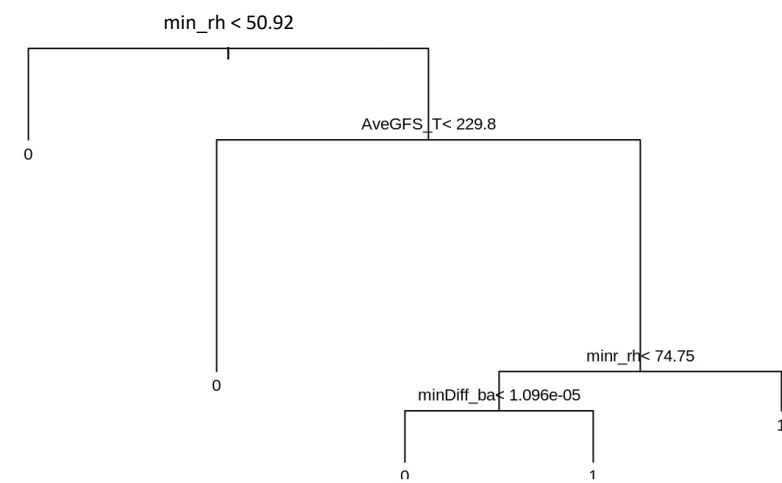
Simple Decision Tree

min_rh < 50.92

AveGFS_T< 229.8

minr_rh< 74.75

minDiff_ba< 1.096e-05

ROC and AUC for ML Models

AUC - SimpleTreeRH = 0.757
AUC - SimpleTreeGeo = 0.722
AUC - SimpleTreeAll = 0.754
AUC - GBM_Geo = 0.842
AUC - GBM_GeoRH = 0.868
AUC - GBM_All = 0.885

model
- GBM_All
- GBM_Geo
- GBM_GeoRH
- SimpleTreeAll
- SimpleTreeGeo
- SimpleTreeRH