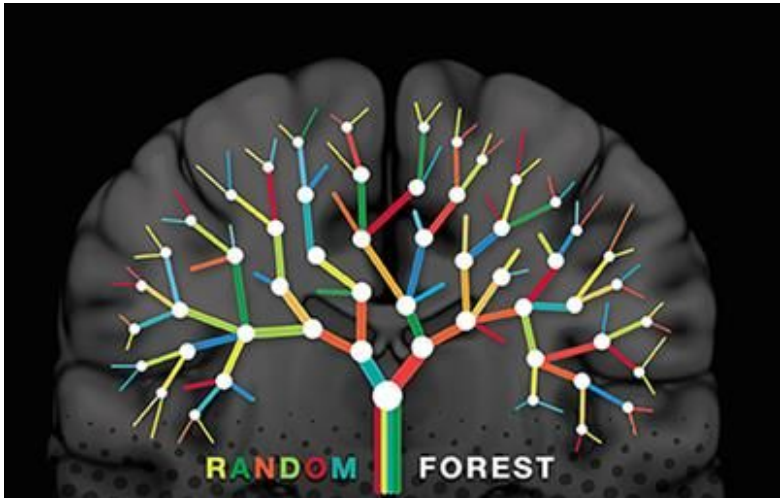


Weather Model

Group H

Dongwei Fu, Joyce Yang, Jun Zhang

4/7/2020

Overview of the Model and Workflow

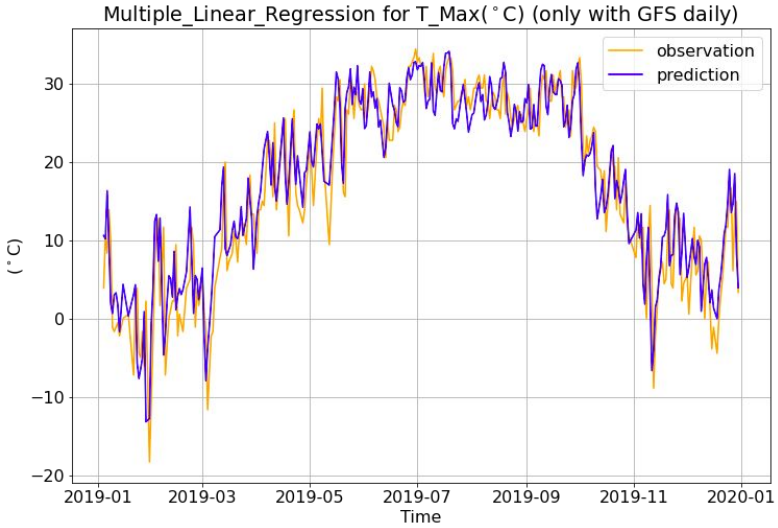
1. **Processing of Observation data**
   a. For KCMI hourly data, sum up the hourly precipitation for each day to get daily precipitation data
   b. Replace the precipitation data in the KCMI daily files with daily precip calculated from hourly data
   c. Other variables (tmax, tmin and wmax) used from daily files directly
   d. Process the KCMI hourly from 6UTC up to 0UTC (not included) prior to the forecast period (shifted a day to better reflect real-world wx-challenge situation), resample to daily on min, mean, max, and sum.
2. **Processing of GFS data**
   a. For GFS surface/profile data, resampled to daily on min, mean, max, and sum.
   b. Replaced 9999/-9999 values with NaN, dropped NaN rows.
   c. Unify the units for variables in Obs and GFS data.

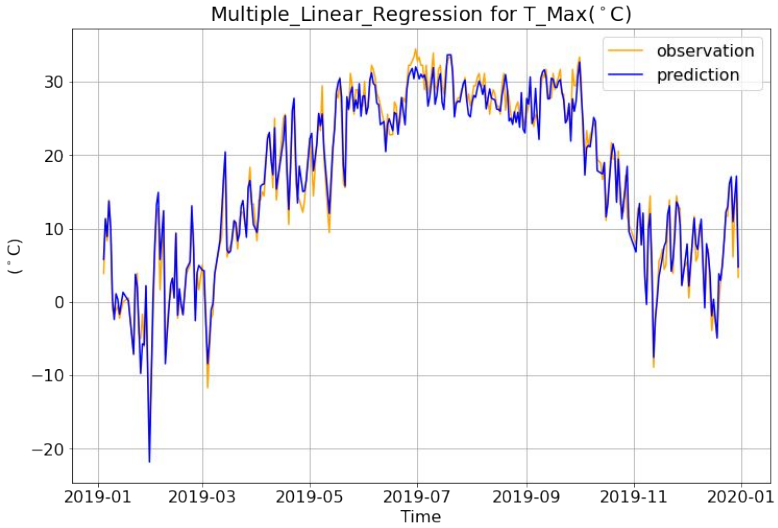
Model Improvement

1. **Selection of features**
    a. Correlation between variables
    b. Feature importance
    c. Manually going through and changing different combinations, taking features in/out

2. **Tuning hyperparameters**
    a. Number of estimators
    b. Maximum number of features

Tmax - MLR





GFS daily only

The Explained Variance: 0.87
The Mean Absolute Error: 3.29 degrees celsius
The Median Absolute Error: 2.52 degrees celsius
The Root Mean Square Error: 4.27 degrees celsius

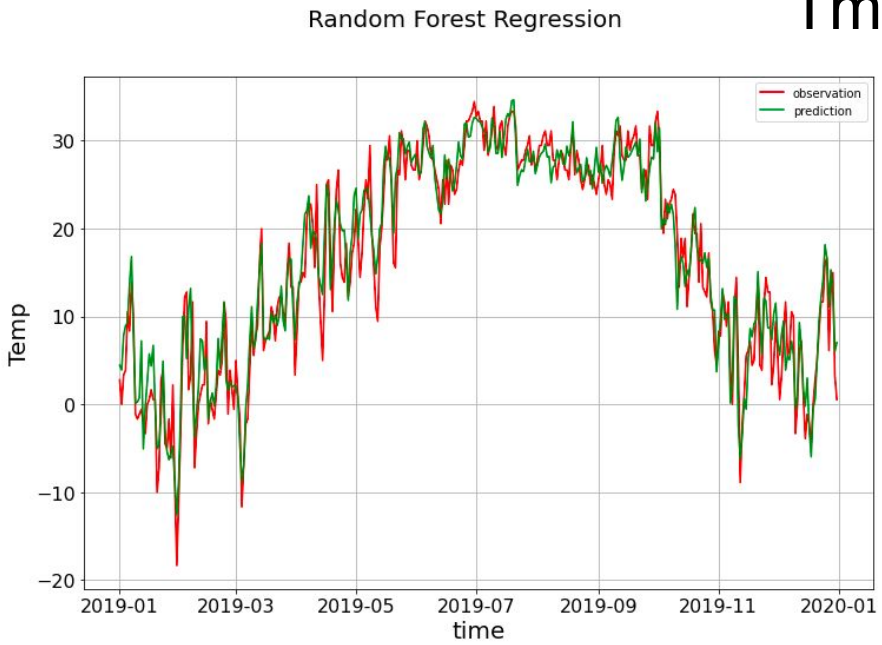
['TMAX_GFS', 'TMIN_GFS', 'WMAX_GFS', 'RTOT_GFS']

GFS daily + vertical + surface + hourly Obs

The Explained Variance: 0.98
The Mean Absolute Error: 1.31 degrees celsius
The Median Absolute Error: 1.10 degrees celsius
The Root Mean Square Error: **1.70** degrees celsius

[all predictors in our dataframe]

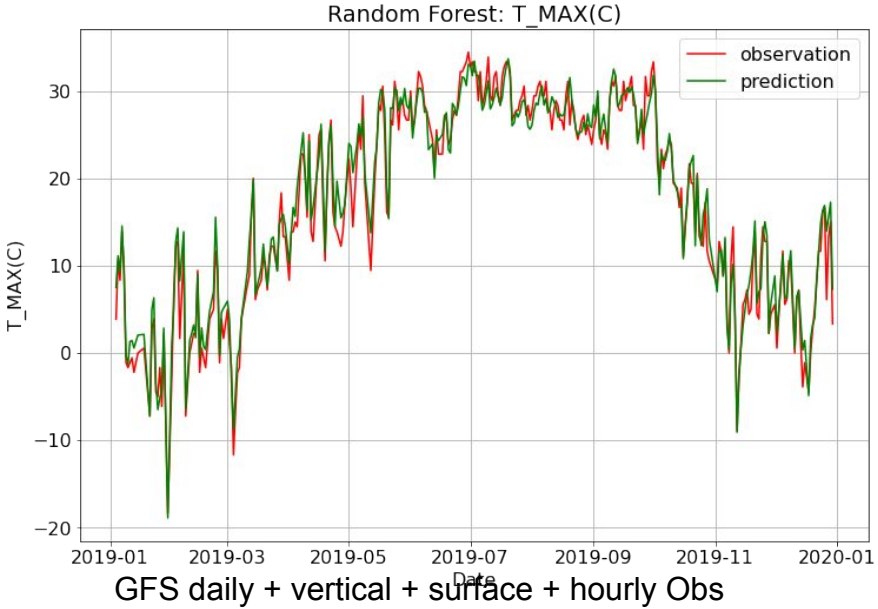
Tmax - RF



GFS daily only

The Explained Variance: 0.93
The Mean Absolute Error: 2.32 degrees celsius
The Median Absolute Error: 1.76 degrees celsius
The Root Mean Square Error: 3.05 degrees celsius

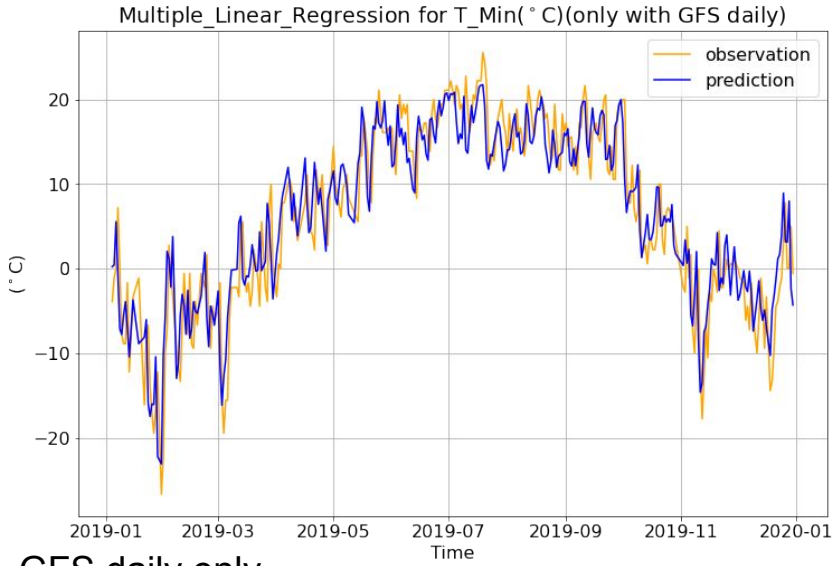
['TMAX_GFS', 'TMIN_GFS', 'WMAX_GFS', 'RTOT_GFS']



GFS daily + vertical + surface + hourly Obs

# of trees = 1000 Max features = 2
Coefficient of determination: 0.97
The Mean Absolute Error: 1.41 degrees celsius
The Median Absolute Error: 1.08 degrees celsius
The Root Mean Square Error: **1.89** degrees celsius
['TMAX_GFS', 'TMIN_GFS', 'TMPC_max', 'TMPC_925_min', 'TMPC_850_min', 'DWPC_925_max', 'DWPC_925_mean', 'tmpc_max']

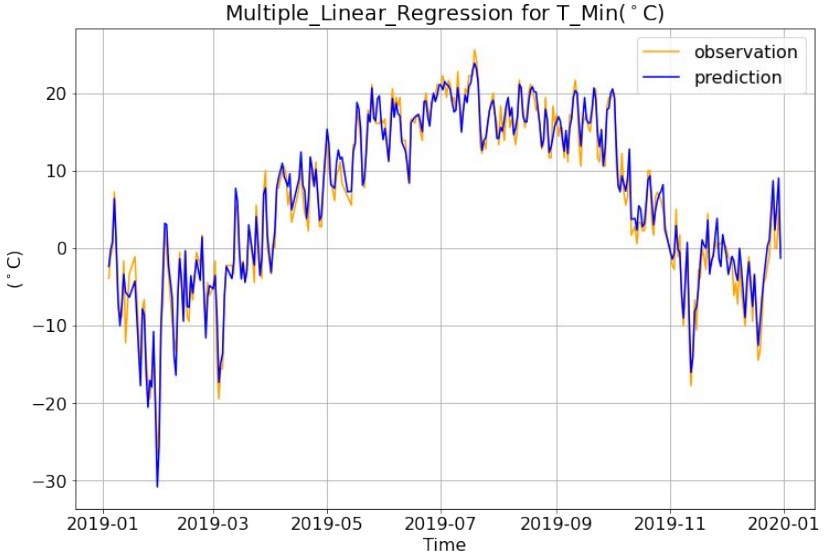
Tmin - MLR



GFS daily only

The Explained Variance: 0.86
The Mean Absolute Error: 3.24 degrees celsius
The Median Absolute Error: 2.82 degrees celsius
The Root Mean Square Error: 4.08 degrees celsius

['TMAX_GFS', 'TMIN_GFS', 'WMAX_GFS', 'RTOT_GFS']

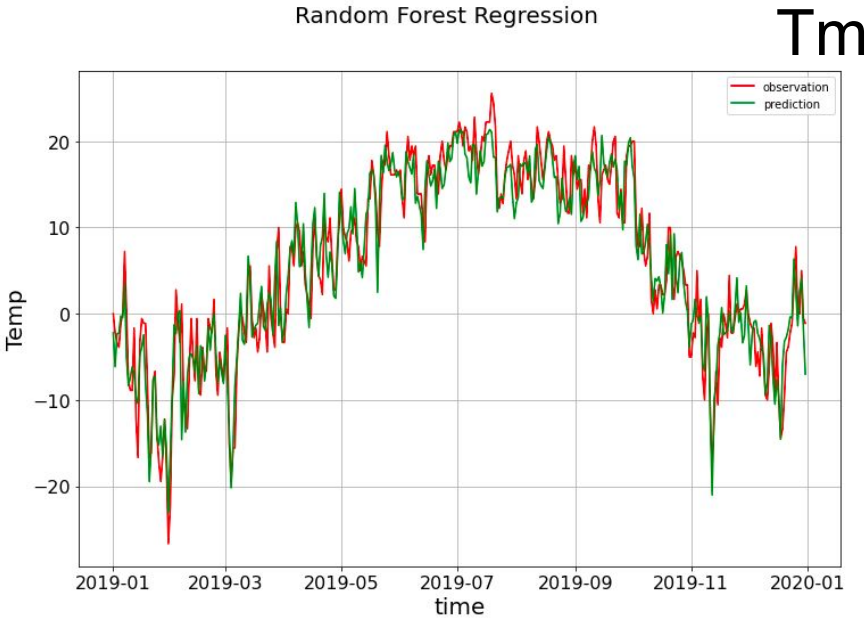


GFS daily + vertical + surface + hourly Obs

The Explained Variance: 0.97
The Mean Absolute Error: 1.37 degrees celsius
The Median Absolute Error: 1.16 degrees celsius
The Root Mean Square Error: **1.73** degrees celsius

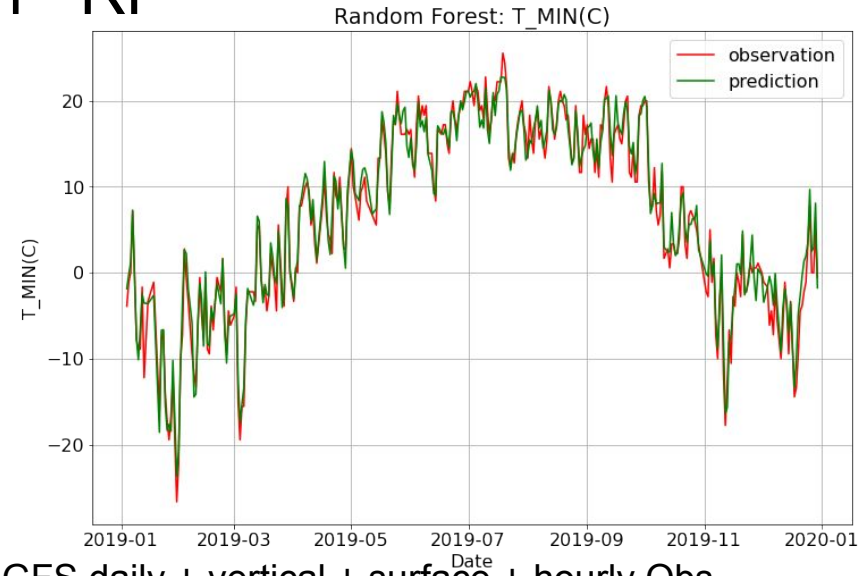
[all predictors in our dataframe]

Tmin - RF



GFS daily only

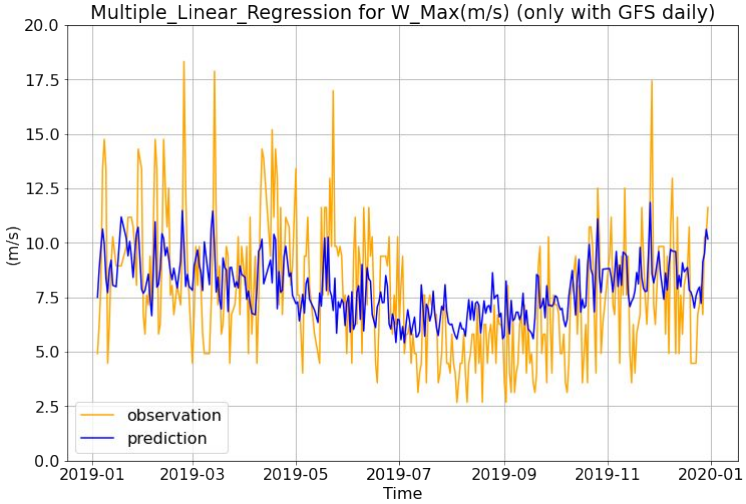
The Explained Variance: 0.92
The Mean Absolute Error: 2.36 degrees celsius
The Median Absolute Error: 1.81 degrees celsius
The Root Mean Square Error: 3.10 degrees celsius
['TMAX_GFS', 'TMIN_GFS', 'WMAX_GFS', 'RTOT_GFS']

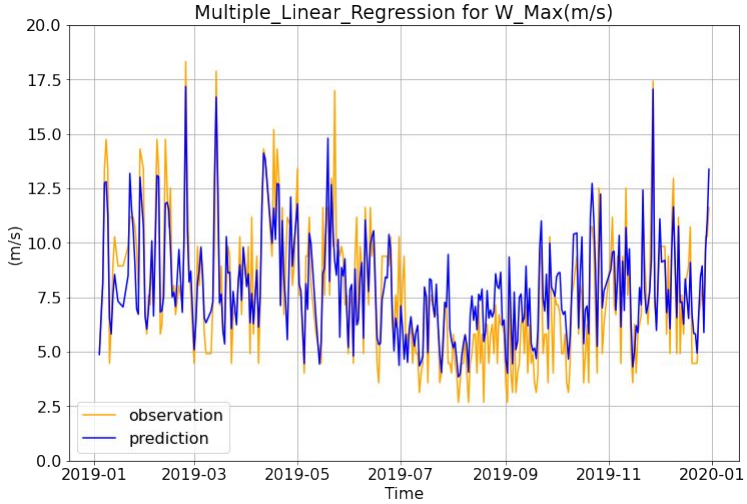


GFS daily + vertical + surface + hourly Obs

# of trees = 1000 Max features = 5
Coefficient of determination: 0.97
The Mean Absolute Error: 1.32 degrees celsius
The Median Absolute Error: 1.11 degrees celsius
The Root Mean Square Error: **1.72** degrees celsius
['TMAX_GFS', 'TMIN_GFS', 'TMPC_min','DWPC_max','TMPC_max',
'TMPC_925_min','TMPC_500_min','TMPC_700_min',
'TMPC_850_min', 'DWPC_925_min','DWPC_925_mean',
'tmpc_min','tmpc_max']

Wmax - MLR





GFS daily only

The Explained Variance: 0.22
The Mean Absolute Error: 2.11 m/s
The Median Absolute Error: 1.86 m/s
The Root Mean Square Error: 2.64 m/s

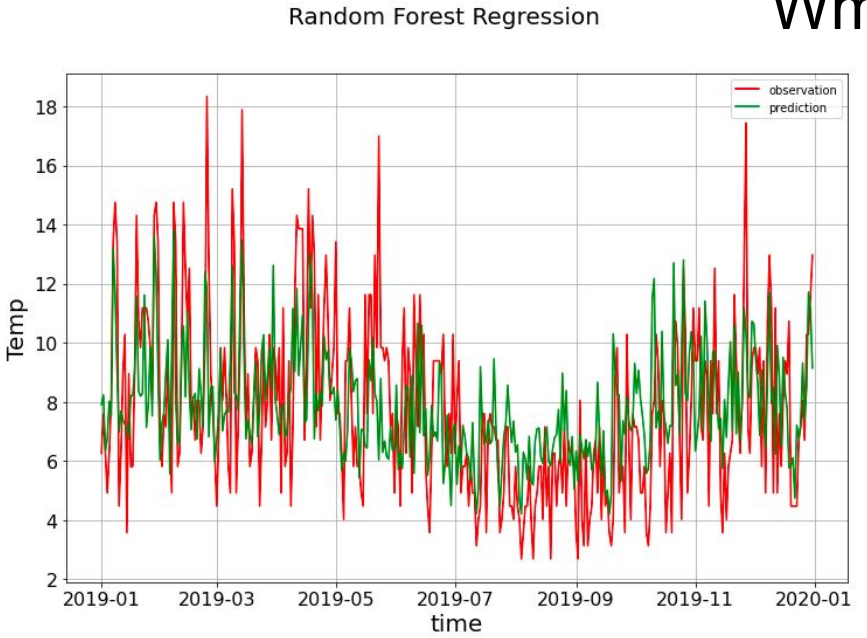
['TMAX_GFS', 'TMIN_GFS', 'WMAX_GFS', 'RTOT_GFS']

GFS daily + vertical + surface + hourly Obs

The Explained Variance: 0.71
The Mean Absolute Error: 1.24 m/s
The Median Absolute Error: 1.07 m/s
The Root Mean Square Error: **1.59** m/s

['WSPD max','WSPD total','WSPD mean','WSPD min','UWND_850_max','UWND_700_max','UWND_925_max','UWND_max','UWND_850_total','UWND_850_mean','UWND_700_total','UWND_700_mean','WMAX','UWND_500_max','VWND_500_max','wspd_max']

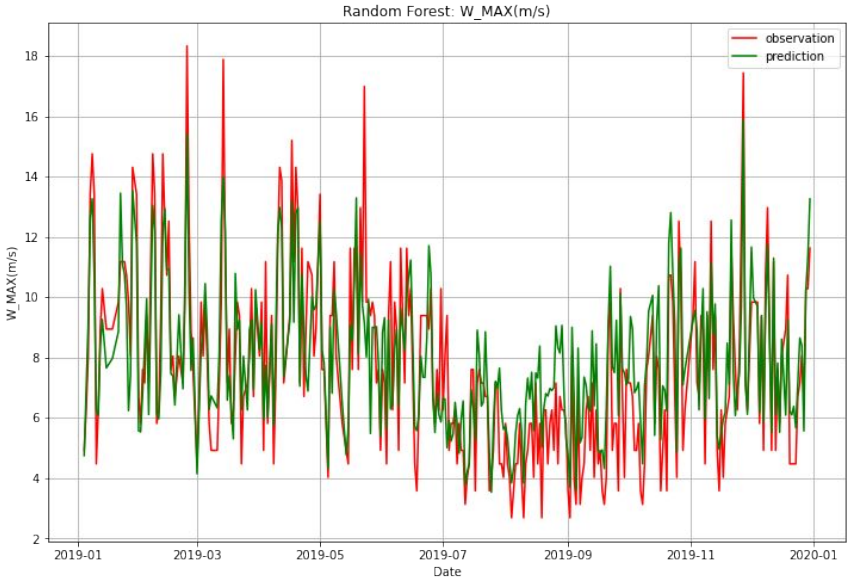
Wmax - RF



GFS daily only

The Explained Variance: 0.33
The Mean Absolute Error: 1.97
The Median Absolute Error: 1.68
The Root Mean Square Error: 2.47 m/s

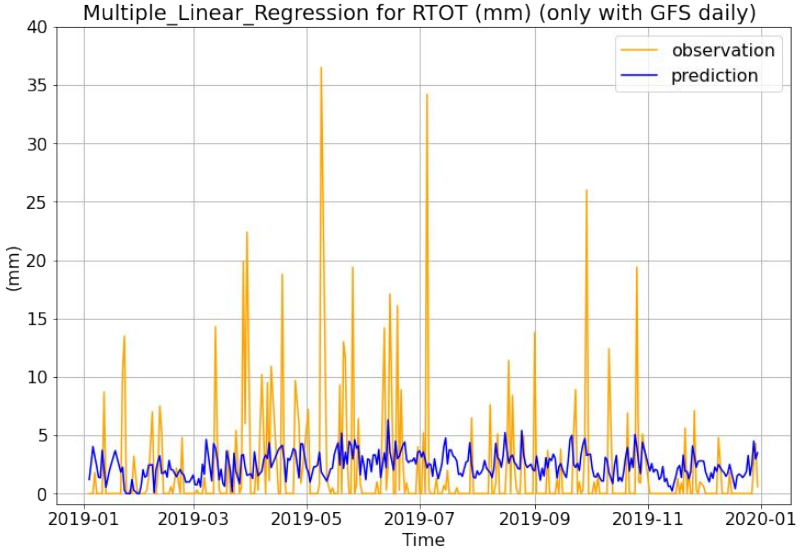
['TMAX_GFS', 'TMIN_GFS', 'WMAX_GFS', 'RTOT_GFS']

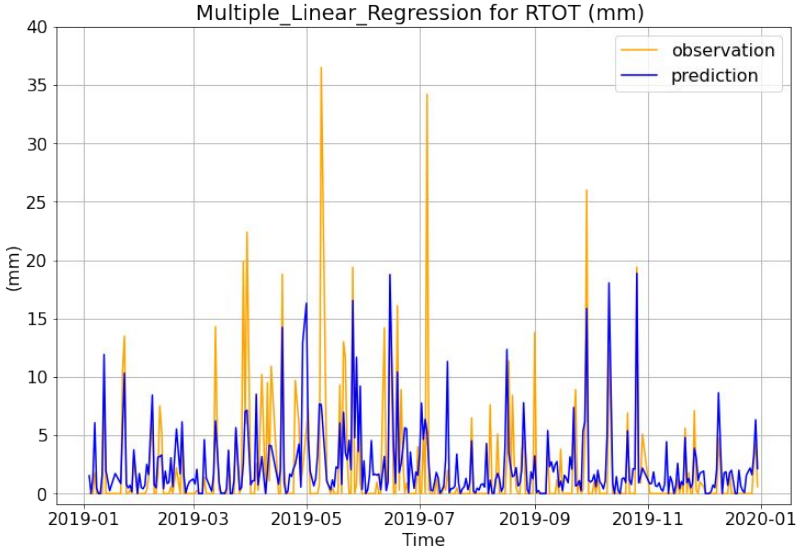


GFS daily + vertical + surface + hourly Obs

# of trees = 1000 Max features = 2
Coefficient of determination: 0.72
The Mean Absolute Error: 1.23 m/s
The Median Absolute Error: 1.02 m/s
The Root Mean Square Error: **1.57** m/s
['WMAX', 'UWND_max', 'WSPD_max', 'WSPD_mean', 'HGHT_500_min', 'HGHT_700_min', 'UWND_850_mean']

Precip - MLR





GFS daily only

The Explained Variance: -0.01
The Mean Absolute Error: 3.08 mm
The Median Absolute Error: 2.13 mm
The Root Mean Square Error: 4.87 mm

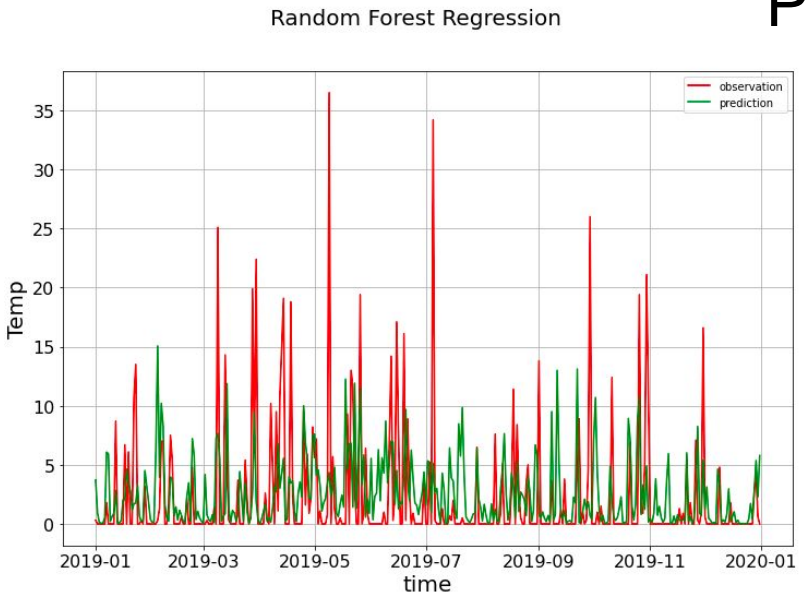
['TMAX_GFS', 'TMIN_GFS', 'WMAX_GFS', 'RTOT_GFS']

GFS daily + vertical + surface + hourly Obs

The Explained Variance: 0.37
The Mean Absolute Error: 2.09 mm
The Median Absolute Error: 1.08 mm
The Root Mean Square Error: **3.87** mm

['PRCP_Sfc_total','PRCP_Sfc_max','HCLD_Sfc_mean',
'VWND_700_max','VWND_500_max','UWND_500_mean']

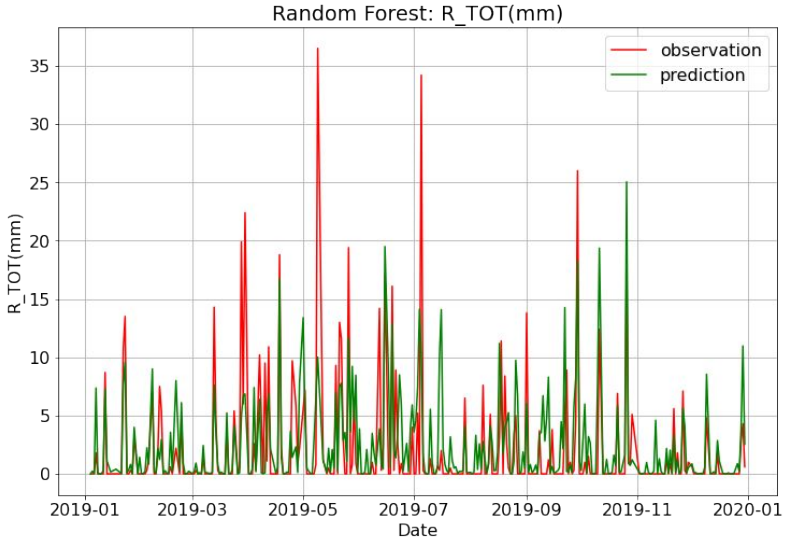
Precip - RF



GFS daily only

The Explained Variance: -0.01
The Mean Absolute Error: 2.91 mm
The Median Absolute Error: 1.26 mm
The Root Mean Square Error: 5.12 mm

['TMAX_GFS', 'TMIN_GFS', 'WMAX_GFS', 'RTOT_GFS']



GFS daily + vertical + surface + hourly Obs

# of trees = 1000 Max features = 1
Coefficient of determination: 0.34
The Mean Absolute Error: 1.93 mm
The Median Absolute Error: 0.48 mm
The Root Mean Square Error: **3.78** mm
['PRCP_max', 'PRCP_mean','DWPC_700_max', 'DWPC_500_min', 'VWND_500_max']

Thank You