1. **src**
   1. **latlon.R**

* Running this first (for some program, e.g. MAIAC\_GRID)
* This is the ranges of lat/lon
  1. **fun.R**
* This includes functions used in all programs
  1. **readHDF.m**
* Reading HDF4 in MATLAB

1. **MAIAC\_GRID**
2. Creating a reference lat/lon grid for the whole project using MAIAC grid
3. Main: maiac\_grid.R
4. Input: input/NA\_LatLon.csv (This should be produced manually)
5. Output: output/maiac\_grid.csv (This grid file is used in all subsequent programs; this is the standard grid file)
6. Run: running directly (Mac/Windows)
7. **Data Preparation**
   1. **MAIAC\_AOD**
      1. **main\_AOD\_CSV.m**
8. Converting original MAIAC HDF (HDF4) files to CSV files
9. Input: /home/jbi6/aura/NorthAmerica\_2000-2016/
10. Output: /home/jbi6/aura/NYS\_MAIAC\_CSV/
11. Run: qsub script\_csv.pbs (Linux)
    * 1. **maiac.script.R**
12. Producing daily MAIAC AOD CSV files in standard grid (IDW Interp)
13. Input: /home/jbi6/aura/NYS\_MAIAC\_CSV/
14. Output: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/MAIAC\_AOD/
15. Run: bash submit.sh (rely on qsub\_script.pbs) (Linux)
16. Notice: the jobs are assigned as [year \* AOD\_type \* cluster\_num]
    1. **NARR**
17. Producing daily NARR weather RData files from NARR monolevel NC files
18. Main: narr.R
19. Input: /home/jbi6/terra/NARR\_ORI/
20. Output
    1. /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/NARR/
    2. Organization: yyyy/varname/yyyyddd\_varname.RData
21. Run: bash submit.sh (Linux)
    1. **MODIS\_CLOUD (MOD06/MYD06)**
       1. **mainCSV.m**
22. Converting original MODIS HDF (HDF4) files into CSV files
23. Input: /home/jbi6/terra/MODIS\_Cloud/.../hdf/
24. Output: /home/jbi6/terra/MODIS\_Cloud/.../csv/
25. Run: qsub script\_csv.pbs (Linux)
    * 1. **modis.script.R**
26. Producing daily MODIS Cloud CSV files in standard grid (IDW Interp)
27. Input: /home/jbi6/terra/MODIS\_Cloud/.../csv/
28. Output: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/MODIS\_Cloud/
29. Run: bash submit.sh (rely on qsub\_script.pbs) (Linux)
30. Notice: the jobs are assigned as [year \* cluster\_num]
    1. **ASTER\_GDEM**
31. Generating DEM file (CSV format) for the research domain
32. Main: main.R
33. Input: /home/jbi6/aura/NYS\_DEM/
34. Output: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/ASTER\_GDEM/
35. Run: qsub qsub\_script
    1. **MODIS\_SNOW (MOD10/MYD10)**
       1. **Snow Testing**
36. Testing the pattern of snow in NYS
37. Main: snow.m
38. Input: Local: test\_data/
39. Output
    1. Local: img/ (image files)
    2. Local: video/ (animation)
       1. **mainCSV**
40. Converting original MODIS HDF (HDF4) files into CSV files
41. Input: /home/jbi6/aura/MODIS\_SNOW\_ORI/.../hdf/
42. Output: /home/jbi6/aura/MODIS\_SNOW\_ORI/.../csv/
43. Run: qsub script\_csv.pbs (Linux)
    * 1. **Modis.script.R**
44. Producing daily MODIS Snow CSV files in standard grid (IDW Interp)
45. Input: /home/jbi6/aura/MODIS\_SNOW\_ORI/.../csv/
46. Output: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/MODIS\_Snow/
47. Run: bash submit.sh (rely on qsub\_script.pbs) (Linux)
48. Notice: the jobs are assigned as [year \* cluster\_num]
    1. **AERONET**
49. Generating a list of AERONET AOT observations (470/550) from difference AERONET site
50. Main: main.R
51. Input: Local: input/
52. Output
    1. Local: output/
    2. Cluster: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/AERONET/
53. Run: in local folder (not cluster), run main.R
54. Notice: ‘site\_info.csv’should be manually made and checked before running the main.R. This CSV file includes the information (i.e. lat/lon, band info, and site name) of each AERONET site, which have the same sequence as the sequence of LEV20 file names in R.
    1. **PM25**
55. Generating a list of PM2.5 ground measurements from ground sites
56. Main: main.R
57. Input
    1. EPA AQS (local): data/AQS/
    2. NAPS (local): data/NAPS/
58. Output
    1. Local: output/
    2. Cluster: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/PM25/
59. Run: in local folder (not cluster), run main.R
    1. **POP**
       1. **readnc.m**
60. Converting the input nc files into csv files (since there is no ‘ncdf4’package in cluster)
61. Main: readnc.m
62. Input: /home/jbi6/terra/POP\_ORI/nc/
63. Output: /home/jbi6/terra/POP\_ORI/
64. Run: qsub script\_readnc (Linux)
    * 1. **pop.R**
65. Generating the LandScan population data
66. Main: pop.R
67. Input: /home/jbi6/terra/POP\_ORI/
68. Output: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/POP/
69. Run: bash submit.sh (Linux)
    * 1. **pop\_plot.R**
70. Plotting the population distribution and checking the population amount
71. Input: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/POP/
    1. **NLDAS** 
       1. **Bash**
          1. **ReorgGRB.sh**
72. Reorganizing the NLDAS files and folders which are downloaded from GES DISC
73. Location: /home/jbi6/aura/NLDAS\_ORI/
    * + 1. **ConvNC.sh**
74. Converting the original \*.grb files into \*.nc files
    1. the bash codes are from Jess, which depends on “/home/jhbelle/.profile”
75. Location: /home/jbi6/aura/NLDAS\_ORI/
    * + 1. **CountNum.sh**
76. Counting how many files are in each folder of NLDAS data
77. Location: /home/jbi6/aura/NLDAS\_ORI/
    * + 1. **Download\_scripts**
78. Scripts used to download NLDAS \*.grb files from GES DISC
79. Location: /home/jbi6/aura/NLDAS\_ORI/download\_scripts/
    * 1. **nldas.R**
80. Generating NLDAS meteorological data
81. Main: nldas.R
82. Input: /home/jbi6/aura/NLDAS\_ORI/data/
83. Output: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/NLDAS/
84. Run: bash submit.sh (Linux)
    1. **GLOBCOVER**
85. Calculating the mode of GRIDCODE for each grid and generating landuse data (csv)
86. Main: globcover.R
87. Input: /home/jbi6/terra/GLOBCOVER\_ORI/grid\_land.csv
88. Output: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/GLOBCOVER/globcover.csv
89. Run: bash submit.sh (Linux)
    1. **NDVI**
90. Converting the original NDVI files (from Keyong Huang) to daily NDVI data within the MAIAC grid
91. Main: ndvi.R
92. Input: /home/jbi6/terra/NDVI\_ORI
93. Output: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/NDVI/
94. Run: bash submit.sh (Linux)
95. **Combine**
96. Combining all parameters into a single (daily) csv file
97. Main: combine.R
98. Input: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/
99. Output: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/Combine/yyyy/
100. Run: bash submit.sh
101. **MI**

* Conduct the multiple imputation

1. **RF**
   1. **rf.R**
2. Gap-filling AOD missing values by random forest
3. Input: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/Combine
   1. YYYYDOY\_combine.RData
4. Output: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/RF
   1. aqua550 & terra550
   2. **YYYYDOY\_RF.RData**
   3. **YYYYDOY\_RF\_MODELPERF.RData** (Model Performace: including rsq, mse, and variable importance measures)
5. Run: bash submit.sh
   1. **push\_time.sh**
6. Automatic submitting a new year's RF Gap-filling when last year's modeling is done
7. Run: bash push\_time.sh -> calling submit\_time.sh $year
   1. **rf\_testing.R**
8. Testing if the fitting ability of RF decreases as choosing less sample size, in order to choose a suitable sample size to speed up the RF fitting
9. The sample sizes tested are 10000, 30000, 50000, and 100000. The prediction results from these sample sizes will be compared to which with complete sample
10. Input: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/Combine
    1. (YYYYDOY\_combine.RData)
11. Output: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/RF\_TEST
    1. (DOY\_SampleSize\_TEST\_RF.RData)
    2. Each output is a result predicted by a certain sample size
12. Run: bash submit\_test.sh
    1. **rf\_testing\_plot.R**
13. To compare the prediction results from the fitting use different sample size (comparison standard is correlation r)
14. Input:
    1. Cluster: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/RF\_TEST
       1. (DOY\_SampleSize\_TEST\_RF.RData)
    2. Local: testdata/
       1. (DOY\_SampleSize\_TEST\_RF.RData)
15. Output
    1. correlation coefficients – R
    2. Scatter plots
16. Run: in local folder (not cluster), run rf\_testing\_plot.R
17. **Modeling**
    1. **Random Forest**
       1. **RF\_Modeling.R**
18. PM2.5 modeling using random forest
19. Input
    1. Gap-filled AOD: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/RF/
    2. Combined data: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/Combine/
20. Output
    1. Combining gap-filled AOD and other parameters
       1. /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/Modeling/PM25\_FIT\_RFMODEL/YYYY/
       2. **YYYY\_COMBINE\_RF.RData**
    2. Random Forest model which will be used to predict PM2.5
       1. /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/Modeling/PM25\_FIT\_RFMODEL/YYYY/
       2. **YYYY\_RFMODEL\_RF.RData**
    3. The fitting results and CV results of RF model
       1. /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/Modeling/PM25\_FIT\_RFMODEL/YYYY/
       2. **YYYY\_RFModel.txt**
21. Run: bash submit\_rf\_model.sh (Linux)
    * 1. **RF\_Pred.R**
22. PM2.5 prediction using random forest
23. Input
    1. Gap-filled AOD: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/RF/
    2. Combined data: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/Combine/
    3. RF model
       1. /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/Modeling/PM25\_FIT\_RFMODEL/YYYY/
       2. **YYYY\_RFMODEL\_RF.RData**
24. Output
    1. /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/Modeling/PM25\_PRED\_RFMODEL/YYYY/
25. Run: bash submit\_rf\_pred.sh (Linux)
26. **Validation**
    1. **AERONET**
27. Comparing gap-filled AOD (RF and MI) with AERONET AOD
28. Main: Combine4AERO.R
    1. Combining MI and RF daily results into single files with available AERONET AOD
    2. Input
       1. MI: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/MI/
       2. RF: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/RF/
    3. Output
       1. MI
          1. /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/Validations/AERONET/
          2. **mi\_combine\_aero.RData**
       2. RF
          1. /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/Validations/AERONET/
          2. **rf\_combine\_aero.RData**
    4. Run: bash submit.sh (Linux)
    5. **PLOT2D**
       1. **Combine4PLOT\_AOD.R**
29. Combining MI and RF daily results to annual AOD averages
30. Input
    1. MI: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/MI/
    2. RF: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/RF/
31. Output
    1. MI
       1. /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/Validations/PLOT2D/PLOTAOD/
       2. **mi\_combine\_plot.RData**
    2. RF
       1. /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/Validations/PLOT2D/PLOTAOD/
       2. **xaotddd\_combine\_[ccc].RData (X is AOT type; DDD is wavelength; CCC is original or gap-filled)**
32. Run: bash submit\_AOD.sh (Linux)
    * 1. **plot2daod.R**
33. Plotting spatial distributions of overall, original, and gap-filled AOD; Do the summary statistics of original and gap-filled AOD
34. Input: data/PLOTAOD/
    1. aaot550\_combine.RData
    2. aaot550\_combine\_ori.RData
    3. aaot550\_combine\_gap.RData
    4. taot550\_combine.RData
    5. taot550\_combine\_ori.RData
    6. taot550\_combine\_gap.RData
35. Output: screen
36. Run: in local folder, run plot2daod.R
    * 1. **Combine4PLOT\_PM25.R**
37. Combining daily PM2.5 predictions into an annual average and 4 seasons’ averages
38. Input: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/Modeling/PM25\_PRED\_RFMODEL/YYYY/
39. Output
    1. /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/Validations/PLOT2D/PLOTPM25/YYYY/
       1. **pm25\_combine\_plot.RData**
       2. **pm25\_combine\_plot\_spring.RData**
       3. **pm25\_combine\_plot\_summer.RData**
       4. **pm25\_combine\_plot\_fall.RData**
       5. **pm25\_combine\_plot\_winter.RData**
40. Run: bash submit\_PM25.sh (Linux)
    * 1. **plot2dpm25.R**
41. Plotting spatial distributions of gap-filled AOD (RF and MI) and modeled PM2.5
42. Input
    1. AOD (copy to local)
       1. data/PLOTAOD/mi\_combine\_plot.RData
       2. data/PLOTAOD/rf\_combine\_plot.RData
    2. PM2.5 (copy to local)
       1. data/PLOTPM25/pm25\_combine\_plot.RData
       2. data/PLOTPM25/pm25\_combine\_plot\_spring.RData
       3. data/PLOTPM25/pm25\_combine\_plot\_summer.RData
       4. data/PLOTPM25/pm25\_combine\_plot\_fall.RData
       5. data/PLOTPM25/pm25\_combine\_plot\_winter.RData
    3. **SummaryStat**
       1. **Missing\_AOD\_CSV**
43. Converting the MAIAC AOD HDF files into CSV files with all available AOD data (including missing data)
44. Main: missing\_aod.m
45. Input: /home/jbi6/aura/NorthAmerica\_2000-2016/
46. Output: /home/jbi6/aura/NYS\_MAIAC\_CSV\_Validation/
47. Run: qsub script (Linux)
    * 1. **Missing\_AOD**
         1. **temporal\_missing.R**
48. Calculating the daily temporal trend of missing AOD caused by cloud and snow covers using original HDF-based CSV data
49. Input: /home/jbi6/aura/NYS\_MAIAC\_CSV\_Validation/
50. Output: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/Validations/SummaryStat/Missing\_AOD
    1. Aqua AOT: **YYYY\_AAOT\_Temporal.RData**
    2. Terra AOT: **YYYY\_TAOT\_Temporal.RData**
51. Run: bash submit\_temp.sh (Linux)
    * + 1. **daily\_missing.R**
52. For checking the results of temp\_missing.R, Calculating the daily trend of missing AOD using daily AOD data (cannot link to a reason of missingness)
53. Input: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/MAIAC\_AOD/2015/
54. Output: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/Validations/SummaryStat/Missing\_AOD/
    1. **Daily\_Missing.RData**
55. Run: bash submit\_daily.sh (Linux)
    * + 1. **plot\_temporal\_missing.R**
56. Plotting temporal trend of missing AOD caused by cloud and snow covers (line graph)
57. Input: data/
    1. Aqua AOT: **YYYY\_AAOT\_Temporal.RData**
    2. Terra AOT: **YYYY\_TAOT\_Temporal.RData**
    3. Daily missing AOT: **Daily\_Missing.RData**

**(for checking temp\_missing.R)**

1. **CaseStudies**
   1. **CloudOnlyAOD**
      1. **rf.R**
2. Only using cloud parameter (cloud\_frac\_day) to fill the missing AOD by random forest
3. Input: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/Combine/
4. Output: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/CaseStudies/CloudOnlyAOD/RF\_CloudOnly/
   1. aqua550 & terra550
   2. **YYYYDOY\_RF.RData**
   3. **YYYYDOY\_RF\_MODELPERF.RData** (Model Performace: including rsq, mse, and variable importance measures)
5. Run: bash submit\_rf.sh (Linux)
   * 1. **rf\_perform.R**
6. Extracting the daily model performance of RF AOD gap-filling, including OOB R2, MSE, snow parameter importance measures (two types) and cloud parameter importance measures (two types)
7. Input: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/CaseStudies/CloudOnlyAOD/RF\_CloudOnly/
8. Output: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/CaseStudies/CloudOnlyAOD/RF\_CloudOnly/Model\_Perform/
   1. **YYYY/model\_perform.RData**
9. Run: bash submit\_perform.sh (Linux)
   * 1. **rf\_perform\_stat.R**
10. Do the statistics about the daily RF gap-filling model performance, such as calculating mean, and quantiles of OOB R2, MSE; summarizing the frequency of variable importance measures of cloud and snow parameters.
11. Input: data/Model\_Perform/model\_perform.RData
12. Output: screen
13. Run: in local folder, run rf\_perform\_stat.R
    * 1. **Combine4PLOT\_AOD.R**
14. Combining RF gap-filled cloud-only and cloud+snow daily AOD to annual AOD averages
15. Input:
    1. Cloud only:

/home/jbi6/terra/MAIAC\_GRID\_OUTPUT/CaseStudies/CloudOnlyAOD/RF\_CloudOnly/

* 1. Cloud+snow:

/home/jbi6/terra/MAIAC\_GRID\_OUTPUT/RF

1. Output: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/CaseStudies/CloudOnlyAOD/PLOT2D/PLOTAOD/
   1. **xaotddd\_combine\_[cloudonly/cloudsnow].RData (X is AOT type; DDD is wavelength)**
2. Run: bash submit\_AOD.sh (Linux)
   * 1. **plot2daod.R**
3. Plotting spatial distributions of cloud-only and cloud+snow gap-filled AOD; Do the summary statistics of cloud-only and cloud+snow gap-filled AOD
4. Input: data/PLOT2D/PLOTAOD/
   1. **aaot550\_combine\_cloudonly.RData**
   2. **aaot550\_combine\_cloudsnow.RData**
   3. **taot550\_combine\_cloudonly.RData**
   4. **taot550\_combine\_cloudsnow.RData**
5. Output: screen
6. Run: in local folder, run plot2daod.R
   1. **WildFire**
      1. **fire.R**
7. Producing temporal trends and spatial distributions of PM2.5 and Gap-filled AOD (Not include original AOD) for a wildfire event occurring in NYS on 5/5/2015
8. Input
   1. PM2.5: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/Modeling/PM25\_PRED\_RFMODEL/2015/
   2. AOD: /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/RF/2015/
9. Output: '/home/jbi6/terra/MAIAC\_GRID\_OUTPUT/CaseStudies/WildFire/'
   1. **pm25\_fire.RData**
   2. **aaot\_fire.RData**
   3. **taot\_fire.RData**
10. Run: Bash submit.sh (Linux)
    1. **SnowMax**
11. Examine the maximum difference of model performances between cloud-only and cloud/snow models (Day 39 in 2015); Comparing the AOD gap-filling results from two models on Day 39
12. Main: snowmax.R
13. Input:
    1. data/
       1. 2015DDD\_RF\_(CldOnly or CldSnw).RData
       2. 2015DDD\_RF\_MODELPERF\_(CldOnly or CldSnw).RData
       3. 2015DDD\_MYD06\_L2.csv (MODIS Cloud)
       4. 2015DDD\_MYD10C1.csv (MODIS Snow)
    2. RF/data/model\_perform.RData (CldSnw Model Performance)
    3. CaseStudies/CloudOnlyAOD/data/Model\_Perform/model\_perform.RData (CldOnly Model Performance)
    4. Validations/SummaryStat/Missing\_AOD/data/2015\_AAOT\_Temporal.RData (Daily AOD Missing Statistics)
14. Output: Screen
15. Run: in local folder, run snowmax.R
    1. **Model\_AOD**

Running different types of random forest models, including a) the model without AOD, b) the model with only original AOD, c) the model with only gap-filled AOD, and d) the model with cloud-gapfilled AOD

* + 1. **RF\_Modeling.R**

1. Random Forest fitting
2. Input
   1. Covariates
      1. /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/Combine/YYYY
   2. Gap-filled AOD
      1. /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/RF/YYYY
   3. Cloud-only gap-filled AOD
      1. /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/CaseStudies/CloudOnlyAOD/RF\_CloudOnly/YYYY
3. Output
   1. Without AOD
      1. /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/CaseStudies/MODEL\_AOD/WithoutAOD/PM25\_FIT\_RFMODEL/YYYY
   2. Original AOD
      1. /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/CaseStudies/MODEL\_AOD/Original/PM25\_FIT\_RFMODEL/YYYY
   3. Gap-filled AOD
      1. /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/CaseStudies/MODEL\_AOD/Gapfilled/PM25\_FIT\_RFMODEL/YYYY
   4. Cloud-only gap-filled AOD
      1. /home/jbi6/terra/MAIAC\_GRID\_OUTPUT/CaseStudies/MODEL\_AOD/CloudOnly/PM25\_FIT\_RFMODEL/YYYY
4. Run: bash submit\_rf\_model.sh (Linux)