Rough Draft Report V1.0

Ryan Erickson

2023-06-02

This is a rough draft of the code and their results.

Libraries

```
# This is for downloading and processing the .nc files from GridMET: https://www.climatologylab.org/gri
# I used the second option '2. Create "wget script" -> might take this out
library(raster)
```

```
## Loading required package: sp
```

```
library(rasterVis)
```

```
## Loading required package: lattice
```

```
library(ncdf4)
library(lattice)
library(stringr)
library(raster)
library(sf)
```

```
## Linking to GEOS 3.9.3, GDAL 3.5.2, PROJ 8.2.1; sf_use_s2() is TRUE
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:raster':
##
##     intersect, select, union

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Special R Document

Here I'm using 'rsource()' to load the R script where I created/cleaned the gauge locations. The results of it are the gauge locations and their resulting mda8 average.

```
source("ozone_krige.R") #need change for mac when applicable
```

```
# Necessary Folders
path_to_cropped_data = "../final_data/"  #need change for mac when applicable
co_data_list = list.files(path_to_cropped_data)
# Random Forest Variables
# Spatial Variables:
```

Some pathing jargon:

Dist to nearest Road

```
getting_folders = grep("co_roads_2019", co_data_list, value = T)
road_shp_file = "co_roads_2019.shp"
path_to_roads = paste0(path_to_cropped_data,getting_folders,"/")
# C = County
# I = Interstate
# M = Common Name
# O = Other
# S = State recognized
# U = U.S.
roads_shp = st_read(paste0(path_to_roads,road_shp_file))
```

```
## Reading layer 'co_roads_2019' from data source
##   'C:\Users\RErickson\Documents\GitHub\ozone_data\final_data\co_roads_2019\co_roads_2019.shp'
##   using driver 'ESRI Shapefile'
## Simple feature collection with 2944 features and 4 fields
## Geometry type: LINESTRING
## Dimension:     XY
## Bounding box:  xmin: -109.0602 ymin: 36.99251 xmax: -102.0417 ymax: 41.00307
## Geodetic CRS:  NAD83
```

```
roads_transformed = st_transform(roads_shp,crs=CRS(prg))
roads_projected = as(roads_transformed, "Spatial")
dist2road = o3_projected$dist2road[1]=round(rgeos::gDistance(roads_projected,o3_projected[1,]),2)
for(i in 2:nrow(o3_projected)){
  dist2road=c(dist2road,round(rgeos::gDistance(roads_projected,o3_projected[i,]),2))
}
o3_projected$dist2road=dist2road
```

Sum of Roads in 500m Buffer

```
add_road_buffer = read_csv("../final_data/roads_in_500m_buffer.csv")[,c(2,4)] %>%
  filter(site_name != "Aspen Park" & site_name != "Evergreen" & site_name !="Welch")
```

```
## Rows: 11 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): site_name
## dbl (3): OID_, FREQUENCY, SUM_SHAPE_Length_1
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
names(add_road_buffer) = c("site_name","road_length")
add_road_buffer$road_length=ifelse(is.na(add_road_buffer$road_length),0,add_road_buffer$road_length)
o3_projected = merge(x=o3_projected,y=add_road_buffer,by="site_name")
```

Elevation

```
elevation_to_add = raster("../final_data/elevation.tiff")
elevation_projected = raster::projectRaster(elevation_to_add, crs=prg)
o3_projected$elev=round(raster::extract(elevation_projected,o3_projected),2)
```

Temporary Data Frame Creation:

```
year_o3 = as.data.frame(o3_projected) %>%
  dplyr::select(c("site_name","elev","dist2road","road_length","lat","long"),everything())

# use to create dataframe of specific months, ex below is summer
summer_o3 = year_o3 %>%
  dplyr::select(contains(c("site_name","lat","long","elev","dist2road","road_length","Apr","May","Jun",
  pivot_longer(cols = contains(c("Apr","May","Jun","Jul","Aug","Sep","Oct")), names_to = "date", values
# preview data
#summer_o3
```

Spatio-Temporal Variables:

```
new_path = "../final_data/Monthly_Averages/"
monthly_path = list.files(paste0(new_path))
max_rh_files=grep("rmax_",monthly_path, value = T)
max_temp_files=grep("tmmx_",monthly_path, value = T)
max_precip_files=grep("pr_",monthly_path, value = T)
renameing_convention = c(paste0(month.abb,".",2017),
                         paste0(month.abb,".",2018),
                         paste0(month.abb,".",2019),
                         paste0(month.abb,".",2020),
```

```
                        paste0(month.abb,".",2021),
                        paste0(month.abb,".",2022))
coordinates(summer_o3) = c('long', 'lat')
proj4string(summer_o3) = CRS(SRS_string = prg)
AE=summer_o3[which(summer_o3$site_name=="Aurora East"),]
BR=summer_o3[which(summer_o3$site_name=="Boulder Reservoir"),]
DC=summer_o3[which(summer_o3$site_name=="Denver - Camp"),]
HR=summer_o3[which(summer_o3$site_name=="Highland Reservoir"),]
LA=summer_o3[which(summer_o3$site_name=="La Casa"),]
NREL=summer_o3[which(summer_o3$site_name=="National Renewable Energy Labs - Nrel"),]
RF=summer_o3[which(summer_o3$site_name=="Rocky Flats-N"),]
WY=summer_o3[which(summer_o3$site_name=="Welby"),]
```

Average Monthly Precipitation

```
max_precip_to_add = stack(paste0(new_path,max_precip_files)) # create stack of raster bricks (each "sta
#plots
# plot(max_precip_to_add$tmmx_2017_monthly_avg_1.1, main="Precipitation for Jan 2017")
max_precip_projected = raster::projectRaster(max_precip_to_add, crs=prg)
names(max_precip_projected) = renameing_convention
#plots: after running names() (above code)
# plot(max_precip_projected$Jan.2017, main="Precipitation for Jan 2017")
summer_max_precip =max_precip_projected[[c(grep("2018", names(max_precip_projected)),
                                           grep("2019", names(max_precip_projected)),
                                           grep("2020", names(max_precip_projected)),
                                           grep("2021", names(max_precip_projected)),
                                           grep("2022", names(max_precip_projected)))]]
summer_max_precip =summer_max_precip[[c(grep("Apr", names(summer_max_precip)),
                                        grep("May", names(summer_max_precip)),
                                        grep("Jun", names(summer_max_precip)),
                                        grep("Jul", names(summer_max_precip)),
                                        grep("Aug", names(summer_max_precip)),
                                        grep("Sep", names(summer_max_precip)),
                                        grep("Oct", names(summer_max_precip)))]]
```

Average Monthly Temperatures

- rounded K to F formula just in case: 1.8*(K-273) + 32

```
max_temperature_to_add = stack(paste0(new_path,max_temp_files)) # create stack of raster bricks (each "
#plots
# plot(max_temperature_to_add$tmmx_2017_monthly_avg_1.1, main="Temperature for Jan 2017")

max_temperature_projected = raster::projectRaster(max_temperature_to_add, crs=prg)
names(max_temperature_projected) = renameing_convention
#plots: after running names() (above code)
# plot(max_temperature_projected$Jan.2017, main="Temperature for Jan 2017")
summer_max_temps =max_temperature_projected[[c(grep("2018", names(max_temperature_projected)),
                                       grep("2019", names(max_temperature_projected)),
                                       grep("2020", names(max_temperature_projected)),
```

```
                                  grep("2021", names(max_temperature_projected)),
                                  grep("2022", names(max_temperature_projected)))]]
summer_max_temps =summer_max_temps[[c(grep("Apr", names(summer_max_temps)),
            grep("May", names(summer_max_temps)),
            grep("Jun", names(summer_max_temps)),
            grep("Jul", names(summer_max_temps)),
            grep("Aug", names(summer_max_temps)),
            grep("Sep", names(summer_max_temps)),
            grep("Oct", names(summer_max_temps)))]]
```

Average Monthly Relative Humidity

```
max_rh_to_add = stack(paste0(new_path,max_rh_files)) # create stack of raster bricks (each "stack" is a
#plots
# plot(max_rh_to_add$tmmx_2017_monthly_avg_1.1, main="Relative Humidity for Jan 2017")
max_rh_projected = raster::projectRaster(max_rh_to_add, crs=prg)
names(max_rh_projected) = renameing_convention
#plots: after running names() (above code)
# plot(max_rh_projected$Jan.2017, main="Relative Humidity for Jan 2017")
summer_max_rh =max_rh_projected[[c(grep("2018", names(max_rh_projected)),
                                        grep("2019", names(max_rh_projected)),
                                        grep("2020", names(max_rh_projected)),
                                        grep("2021", names(max_rh_projected)),
                                        grep("2022", names(max_rh_projected)))]]
summer_max_rh =summer_max_rh[[c(grep("Apr", names(summer_max_rh)),
                                      grep("May", names(summer_max_rh)),
                                      grep("Jun", names(summer_max_rh)),
                                      grep("Jul", names(summer_max_rh)),
                                      grep("Aug", names(summer_max_rh)),
                                      grep("Sep", names(summer_max_rh)),
                                      grep("Oct", names(summer_max_rh)))]]
```

Exposure Assignment

```
for(i in 1:nrow(AE)) {
  AE$tmax[i] = raster::extract(summer_max_temps[[i]],AE[i,])
  AE$rhmax[i] = raster::extract(summer_max_rh[[i]],AE[i,])
  AE$pmax[i] = raster::extract(summer_max_precip[[i]],AE[i,])
}
for(i in 1:nrow(BR)) {
  BR$tmax[i] = raster::extract(summer_max_temps[[i]],BR[i,])
  BR$rhmax[i] = raster::extract(summer_max_rh[[i]],BR[i,])
  BR$pmax[i] = raster::extract(summer_max_precip[[i]],BR[i,])
}
for(i in 1:nrow(DC)) {
  DC$tmax[i] = raster::extract(summer_max_temps[[i]],DC[i,])
  DC$rhmax[i] = raster::extract(summer_max_rh[[i]],DC[i,])
  DC$pmax[i] = raster::extract(summer_max_precip[[i]],DC[i,])
}
```

```
for(i in 1:nrow(HR)) {
  HR$tmax[i] = raster::extract(summer_max_temps[[i]],HR[i,])
  HR$rhmax[i] = raster::extract(summer_max_rh[[i]],HR[i,])
  HR$pmax[i] = raster::extract(summer_max_precip[[i]],HR[i,])
}
for(i in 1:nrow(LA)) {
  LA$tmax[i] = raster::extract(summer_max_temps[[i]],LA[i,])
  LA$rhmax[i] = raster::extract(summer_max_rh[[i]],LA[i,])
  LA$pmax[i] = raster::extract(summer_max_precip[[i]],LA[i,])
}
for(i in 1:nrow(NREL)) {
  NREL$tmax[i] = raster::extract(summer_max_temps[[i]],NREL[i,])
  NREL$rhmax[i] = raster::extract(summer_max_rh[[i]],NREL[i,])
  NREL$pmax[i] = raster::extract(summer_max_precip[[i]],NREL[i,])
}
for(i in 1:nrow(RF)) {
  RF$tmax[i] = raster::extract(summer_max_temps[[i]],RF[i,])
  RF$rhmax[i] = raster::extract(summer_max_rh[[i]],RF[i,])
  RF$pmax[i] = raster::extract(summer_max_precip[[i]],RF[i,])
}
for(i in 1:nrow(WY)) {
  WY$tmax[i] = raster::extract(summer_max_temps[[i]],WY[i,])
  WY$rhmax[i] = raster::extract(summer_max_rh[[i]],WY[i,])
  WY$pmax[i] = raster::extract(summer_max_precip[[i]],WY[i,])
}

rbind(AE,BR,DC,HR,LA,NREL,RF,WY)
```

```
## class       : SpatialPointsDataFrame
## features    : 272
## extent      : 481219.8, 536954.6, 4379800, 4435552  (xmin, xmax, ymin, ymax)
## crs         : +proj=utm +zone=13 +datum=WGS84 +units=m +no_defs
## variables   : 9
## names       :   site_name,    elev, dist2road,      road_length,     date,              mda8,
## min values  : Aurora East, 1592.47,     68.83,                0, Apr.2018, 18.4914838709677, 286.2353
## max values  :       Welby, 1793.14,  11202.39, 3251.26803095015, Sep.2022, 62.7419677419355, 307.0410
```

```
ggs = as.data.frame(rbind(AE,BR,DC,HR,LA,NREL,RF,WY))
ggs %>%
  group_by(site_name) %>%
  slice_head(n=1)
```

```
## # A tibble: 8 x 11
## # Groups:   site_name [8]
##   site_name       lat   long  elev dist2road road_length date   mda8  tmax rhmax
##   <chr>         <dbl>  <dbl> <dbl>     <dbl>       <dbl> <chr> <dbl> <dbl> <dbl>
## 1 Aurora East  4.39e6 5.37e5 1793.   11202.           0  Apr.~  45.4  290.  64.1
## 2 Boulder Res~ 4.44e6 4.81e5 1594.     545.           0  Apr.~  43.2  290.  66.0
## 3 Denver - Ca~ 4.40e6 5.01e5 1609.    1175.           0  Apr.~  37.6  291.  62.7
## 4 Highland Re~ 4.38e6 5.04e5 1746.     281.        3251. Apr.~  45.9  290.  62.5
## 5 La Casa      4.40e6 5.00e5 1609.     413         1232. Apr.~  37.2  291.  62.7
## 6 National Re~ 4.40e6 4.85e5 1767.    1431.           0  Apr.~  43.7  289.  63.7
## 7 Rocky Flats~ 4.42e6 4.84e5 1728.      68.8       2405. Apr.~  47.4  290.  61.6
```

```
## 8 Welby        4.41e6 5.04e5 1592.     818.           0  Apr.~  34.4  291.  64.1
## # i 1 more variable: pmax <dbl>
```

Still Need:

- yearly NDVI: Average NDVI 500m buffer - need buffer data is ready
- Temporal Variables
- monthly dummy variable
- yearly dummy variable
- Monthly mode of wind direction -> omitting this but not deleting because I may come back to it
- RF Model, “Leave one out” cross validation