HOMEWORK SPATIAL ANALYSIS

NOTES:

- Please submit your homework via email and check the submission instructions.
- There will be two submissions for this homework: one for report in Rmd and the other for code.
- Provide brief introduction of the data (describe variables, provide simple exploratory analysis) as well as specific comments of your choices during the statistical analysis
- Please keep in mind the collaboration policy. If you discuss questions with others you must write their names on your submission, and if you use any outside resources you must reference them. Do not look at each others' writeups, including code.

Please complete the following:

PROBLEM 1

Here we analyze ozone (03) data. The dataset "**pollution.csv**" contains six (6) variables (columns 1-6). The outcome variable Y is given by column 6. The first column provides information on the *year*, the third and fourth contain Lat and Long information and the fifth the type of pollutant. Use the following packages: library(geoR), library(ggplot2), library(maps), library(viridis)

a) Given the full dataset, build the data matrix by selecting the "year=2014" (first column) and "pollutant=O3" (fifth column). This selection leads to the data matrix (call this matrix "dat") of dimension (1255 × 6). Multiply by 100 the last column so the data Y are easier to visualize:

```
Y <- 100*dat[,6] .
```

Define also the following matrix of coordinates

```
s <- as.matrix(dat[,3:4])
```

is.na()?

b) Plot the histogram of the data together with a density plot of the data

c) Build a "data.frame" containing coordinates and spatial data, build the "as.geodata" object and give a plot of the sites, coordinates Vs data and 3D scatter of the data (provide comments)

d) Fit a (polynomial) trend function (you can choose the order) and provide a plot of the fitted surface; provide a variogram plot on the spatial residuals ("res_trend"). To build the variogram first generate the following vector of distances "dist" by using the command:

```
dist <- seq(0,10,length=30)
```

then, obtain the variogram values by setting

variog4

```
vg <- variog(coords = put coordinates here, data = put trend residuals here, uvec = dist)
```

and plot the sample variogram (provide comments of the results);

e) Try to fit the sample variogram by eye by choosing (guessing) the values of the "partial sill" (**sig2**), "nugget"(**tau2**) and "range" (**phi**) parameters. You can also try to build the function

```
vg_fit <- (sig2 + tau2) - sig2*exp(-dist/phi)
```

and plot it onto the sample variogram.

f) Build a regular grid where to make your spatial prediction by using the following coordinates:

```
X0 <- seq(-135,-65,0.1)
Y0 <- seq(25,50,0.1)
```

and use the command "expand.grid" to obtain the final prediction grid

```
s0  <- as.matrix(expand.grid(X0,Y0))
```

g) Use the following commands to map the sites:

```
map("usa")
points(s0,pch=19,cex=0.1)
map("usa",add=TRUE,col="white")
map("state","north carolina")
points(s0,pch=19,cex=0.1)
inusa <- map.where("usa",s0[,1],s0[,2])
s0    <- s0[!is.na(inusa),]
map("usa")
points(s0)
map("state","north carolina")
points(s0,pch=19,cex=0.1)
```

h) Use the function "likfit" to estimate the model parameters with the following setting:

- ini.cov.pars=c(0.1,2), nugget = 0.1
- trend= "….." this choice depends on the trend function estimated above
- cov.model="exponential"

Show the estimated parameters and provide comments on their values.

i) Use the functions "krige.control " and "krige.conv" to obtain kriging predictions on the grid (put the output in "kriging");

l) use the following functions to plot the data and kriging predictions:

```
df <- data.frame(long=s[,1],lat=s[,2],Y=Y)
ggplot(df, aes(long, lat)) +
borders("state") +
geom_point(aes(colour = Y)) +
scale_colour_gradientn(colours = viridis(10)) +
xlab("")+ylab("")+labs(title="Ozone (ppm), 2014")+
coord_fixed()

# Plot Kriging predictions
df <- data.frame(long=s0[,1],lat=s0[,2],Y=kriging$pred)
ggplot(df, aes(long, lat)) +
borders("state") +
geom_raster(aes(fill = Y)) +
scale_fill_gradientn(colours = viridis(10))+
xlab("")+ylab("")+labs(title="Predicted ozone (ppm), 2014")+
coord_fixed()
```

m) plot the map of the variance kriging prediction errors