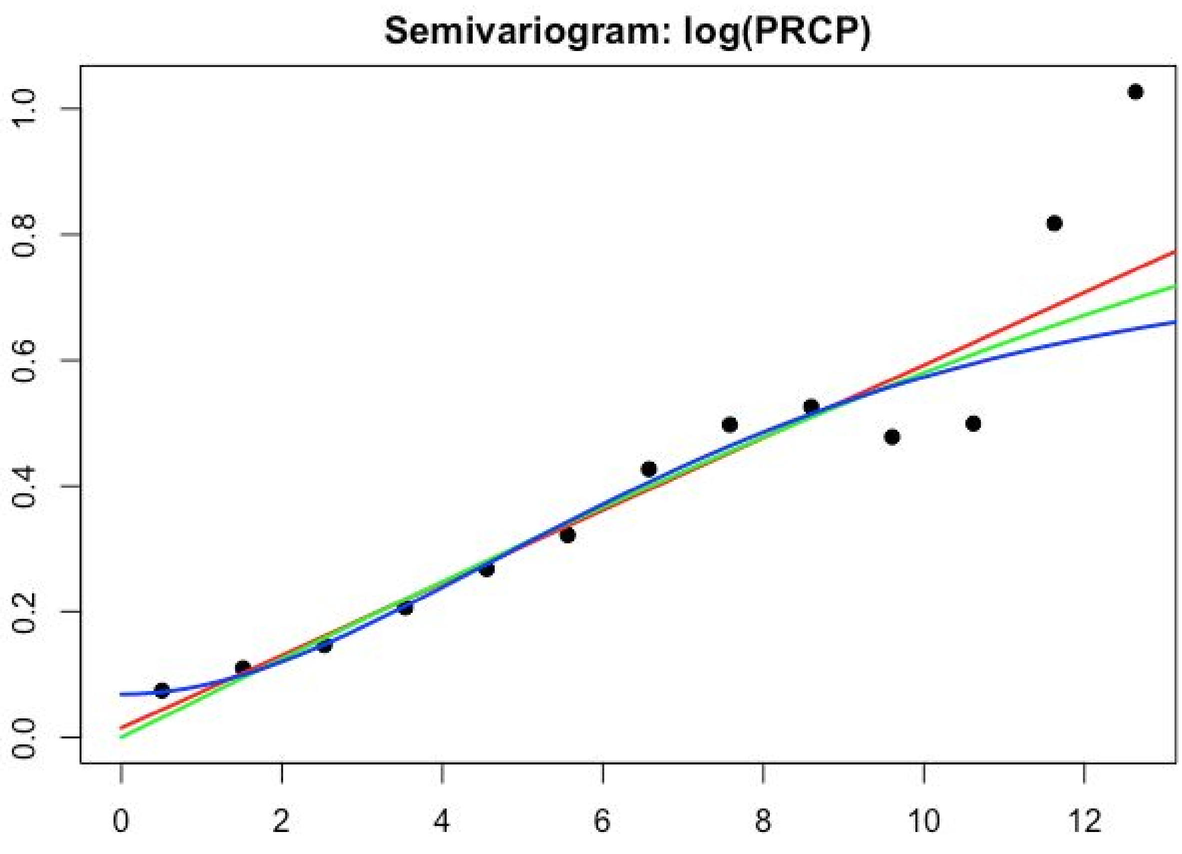
Patrick Tinsley
ACMS 60855 HW #4
Due Date: 4/10

Problem 1: Fit a linear, spherical and Matérn model on this data: which one would be your final choice and why?

See appendix for R code.

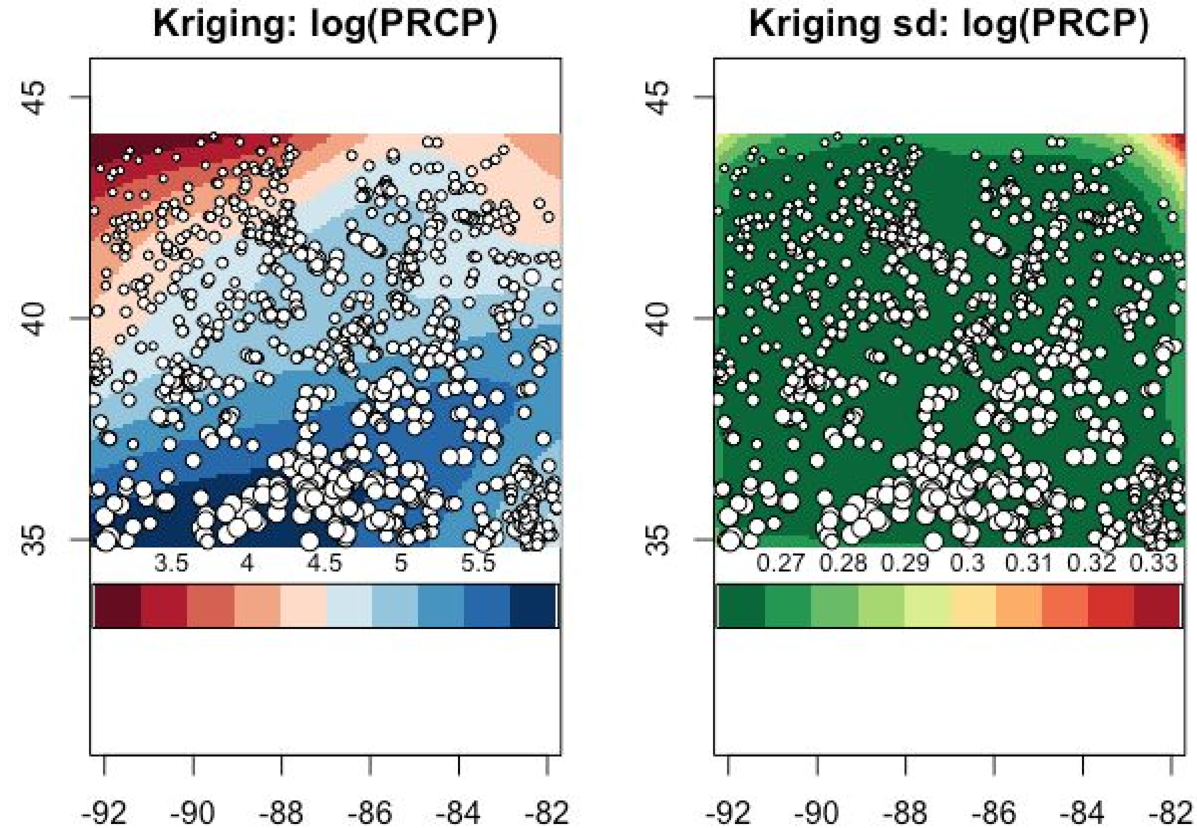
After fitting the three models, we choose the Matérn model for this data set. As seen in the graph below, the chosen model fits the data the closest. Our choice further reinforced by the fact that the Matérn model results in the smallest result for the minimal value of the loss function.



Problem 2: Perform kriging with your chosen model in the previous point, and show a map of the estimated values and the kriging standard deviation. Comment on the results.

See appendix for R code.

The maps of the estimated values and standard deviations can be seen on the next page.

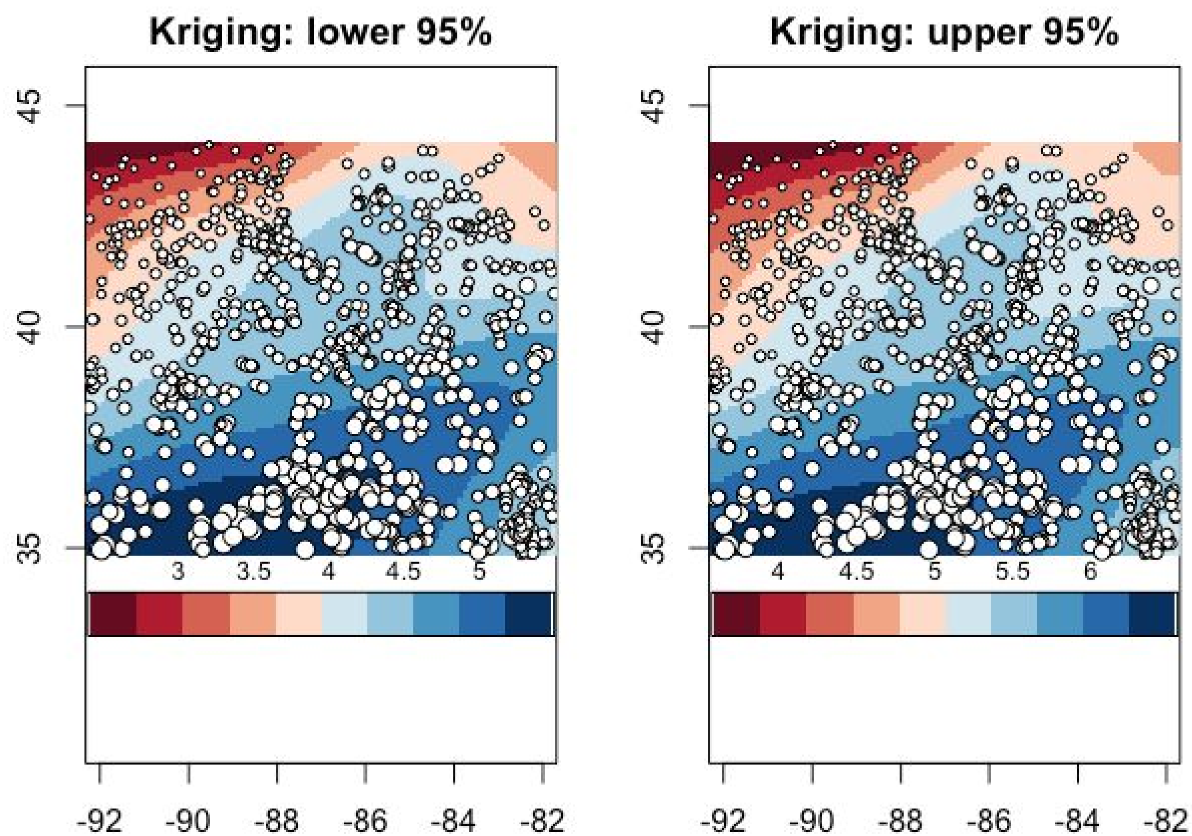


On the left, we see that kriging produced results in line with our data-based intuition. Thinking back to HW3, we observe that the kriging map also suggests heavier rainfall in the more southern regions. Though the units above are log-scale, we can exponentiate the legend values to convert back to the data set's original units. After doing so, we see the bluer regions receive around ~148mm (exp(5)=148.41315) while the red receive ~56mm (exp(4)=55.5981500331). This is visible in our color palette choice, which has darker blue representing more rainfall and red representing less. On the right, we see that the standard deviation of the kriging is rather uniform across the map, save the top-right corner (this could be due to lack of data in that region). Thankfully, the majority of the map is green, which I chose to represent a smaller deviation, which reassures us that the procedure is very precise; the standard deviation hovers around 1.3mm (exp(.26)=1.2969300867).

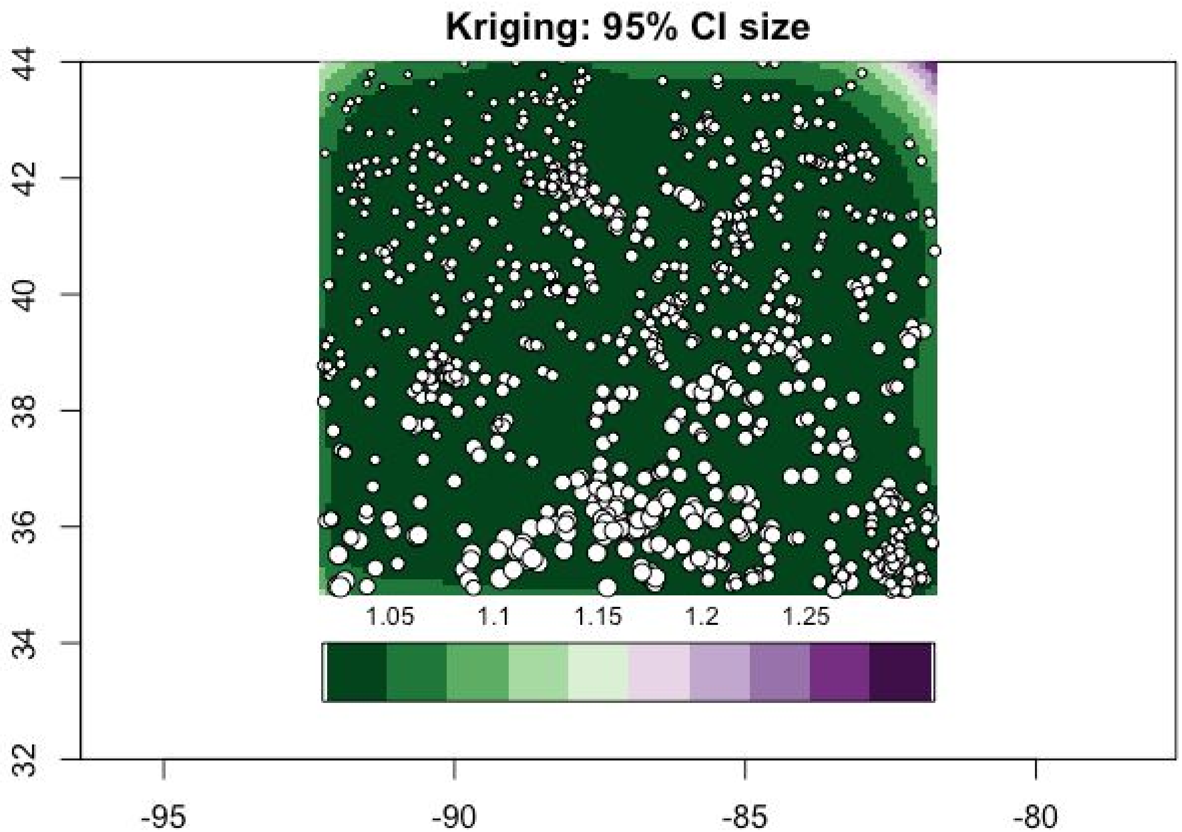
Problem 3: Compute and plot the lower and upper bounds of a 95% confidence interval for every point. Comment on the results and provide an interpretation of the intervals.

See appendix for R code.

The maps of the lower and upper bounds of a 95% confidence interval for every point can be seen on the next page.



Since the general kriging patterns/color trends across both plots above are very similar, what is of most importance for this problem is the scale of the z-axes and the size of the interval at each point. To better address the question, I plotted the interval size at each point in a separate plot on the next page. The graph suggests that the interval sizes are also fairly consistent, just as was the case with standard deviation earlier. According to the plot, z-axis and legend values, the interval sizes hover around ~2.9mm (exp(1.05)=2.8576511181), which is very narrow. In terms of interpretation, we can say that we are 95% confident that the *true* value of the rainfall at any one point on the map will fall between the lower value on the left graph and the upper value on the right graph; it is of great import to note that this is not a family-wise interval. Again, the color palette was chosen to reflect narrow confidence as good (green) and wider intervals as worse (purple).



Problem 4: Would you provide the kriging map to the general public as an official estimate of rainfall in the Midwest for February 2018?

Personally, I would not provide the kriging map to the general public as an official estimate. The first reason for my hesitation comes from experience living in the Midwest for the past five years; I have seen firsthand plenty of rainfall throughout the Midwest, and 3mm is simply unbelievable as a 95% confidence interval size. Secondly, I am not very convinced by the Matérn model fit to the variogram. Though it outperformed the linear and spherical alternatives, I would definitely question the goodness-of-fit for the chosen model; experimenting with an exponential or potentially hybrid model might be a relevant future effort. Thirdly, given the size of our original dataset (1,000 observations), our prediction grid seems a bit too fine with 10,000 cells; if we see consistent results with different values for grid coarseness, I would be more inclined to share these findings. Finally, as is always the case, we would improve our results with more data collection, which almost always leads to better, more reliable findings.

Appendix

```
rm(list=ls())

library(geoR)
library(sp)
library(fields)

df = read.csv('precip.csv')
head(df)

coordinates(df) = ~LONGITUDE+LATITUDE
bubble(df, 'PRCP', col='blue', maxsize=1.5, pch=19, fill=T, main = 'Precipitation')

df.geoR = as.geodata(df,data.col=1)

lz.geoR = df.geoR
lz.geoR$data = log(df.geoR$data)
lz.v = variog(lz.geoR, max.dist=20)

par(mfcol=c(1,1))
plot(lz.v, pch=19)

## fit variogram models
lz.fit.lin = variofit(lz.v, ini=c(0.8,1), cov.model='linear', fix.nugget = F,
nugget=0, kap=2.5)
lz.fit.sph = variofit(lz.v, ini=c(0.8,1), cov.model='spherical', fix.nugget = F,
nugget=0, kap=2.5)
lz.fit.mat = variofit(lz.v, ini=c(0.8,1), cov.model='matern', fix.nugget = F,
nugget=0, kap=2.5)

lz.fit.lin$value
lz.fit.sph$value
lz.fit.mat$value # we choose matern because minimal value of loss function

## plot for visual inspection
plot(lz.v, pch=19, main='Semivariogram: log(PRCP)')
lines(lz.fit.lin, col='red', lwd=2)
lines(lz.fit.sph, col='green', lwd=2)
lines(lz.fit.mat, col='blue', lwd=2)

## create prediction grid
x <- seq(min(lz.geoR$coords[,1]), max(lz.geoR$coords[,1]), length.out = 100)
y <- seq(min(lz.geoR$coords[,2]), max(lz.geoR$coords[,2]), length.out = 100)
gr <- expand.grid(x = x, y = y)

## actual kriging
```

```
kc = krige.control(type='ok', obj.mod = lz.fit.mat)
sk = krige.conv(lz.geoR, krige=kc, loc=gr)

## plotting the kriging maps
par(mfrow=c(1,2))
image(sk,
      x.leg = c(min(df.geoR$coords[,1]), max(df.geoR$coords[,1])), y.leg = c(33, 34),
      ylim = c(32, 44),
      col=brewer.pal(10, "RdBu"),
      main="Kriging: log(PRCP)")
points(df.geoR, add=T)

## plotting kriging standard deviations
image(sk, val = sqrt(sk$krige.var),
      x.leg = c(min(df.geoR$coords[,1]), max(df.geoR$coords[,1])), y.leg = c(33, 34),
      ylim = c(32, 44),
      col=rev(brewer.pal(10, "RdYlGn")),
      main="Kriging sd: log(PRCP)")
points(df.geoR,pch=,add=T)

par(mfrow=c(1,2))
## lower 95% confidence interval
image(sk, val = sk$predict-1.96*sqrt(sk$krige.var),
      x.leg = c(min(df.geoR$coords[,1]), max(df.geoR$coords[,1])), y.leg = c(33, 34),
      ylim = c(32, 44),
      col=brewer.pal(10, "RdBu"),
      main="Kriging: lower 95%")
points(df.geoR,pch=,add=T)

## upper 95% confidence interval
image(sk, val = sk$predict+1.96*sqrt(sk$krige.var),
      x.leg = c(min(df.geoR$coords[,1]), max(df.geoR$coords[,1])), y.leg = c(33, 34),
      ylim = c(32, 44),
      col=brewer.pal(10, "RdBu"),
      main = "Kriging: upper 95%")
points(df.geoR,pch=,add=T)

## interval size
par(mfrow=c(1,1))
image(sk, val = sk$predict+1.96*sqrt(sk$krige.var) -
(sk$predict-1.96*sqrt(sk$krige.var)),
      x.leg = c(min(df.geoR$coords[,1]), max(df.geoR$coords[,1])), y.leg = c(33, 34),
      ylim = c(32, 44),
      col=rev(brewer.pal(10, "PRGn")),
      main = "Kriging: 95% CI size")
points(df.geoR,pch=,add=T)
```