ST533/433 Homework: 3

September 16, 2020

1) Comparison of four models:

	Exp, Nugget	Exp, No Nugget	Matern, Nugget	Matern, No Nugget
AIC	3927.62	3993.36	3929.62	3995.36
BIC	3957.19	4018.00	3964.12	4024.93
MSE	2.34	2.31	2.60	2.82
MAD	0.53	0.55	0.58	0.63
SD	1.49	1.44	1.57	1.68
COV	95.40	95.40	94.81	93.93
COR	0.94	0.94	0.93	0.93

We choose Model 1: Exponential with Nugget effect because of the overall less AIC, BIC values ans well as better MSE, MAD and Coverage values than Models 3 and 4. Further, a no nugget effect spikes the BIC.

Different models can be chosen as per individual justifications.

2) Final Model

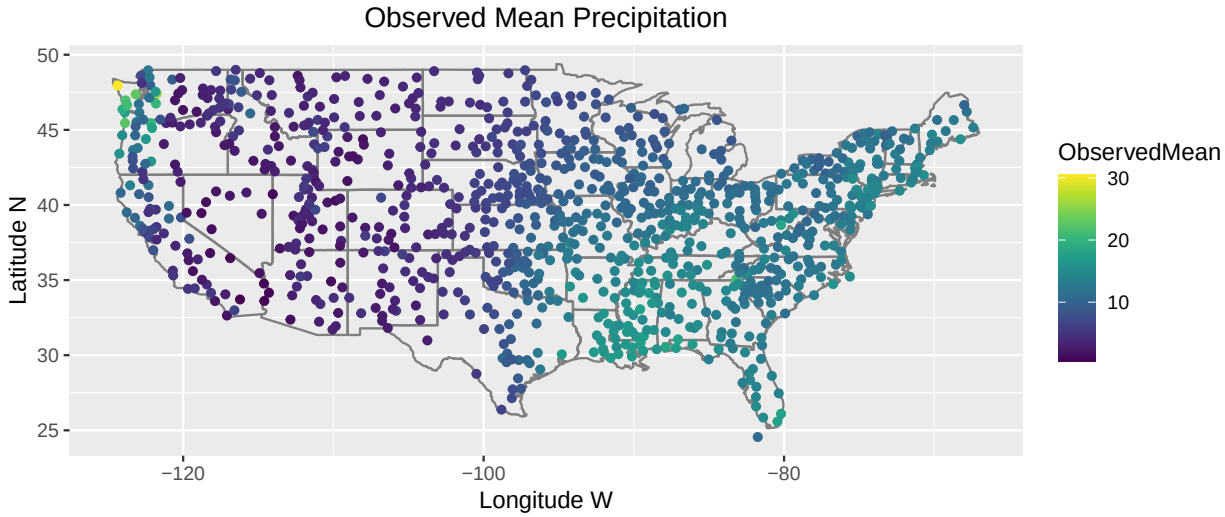
$Y = \beta_0 - \beta_1 X_1 + \beta_2 X_2$, where Y is daily mean precipitation, X_1 is Longitude and X_2 is Latitude.

MLE estimates from Model 1: $\hat{\beta}_0 = 2.44$, $\hat{\beta}_1 = 0.093$, $\hat{\beta}_2 = 0.035$ $\tau^2 = 0.913$, $\sigma^2 = 51.7$, $\phi = 31.44$

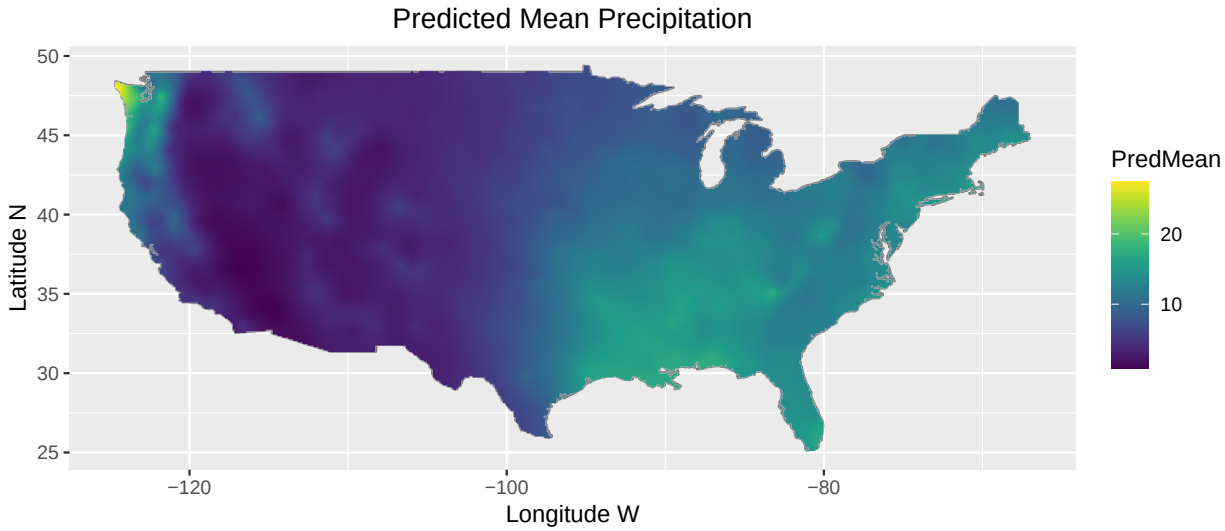
Interpretation: Given a position with coordinates (0,0) the mean daily precipitation is 2.44 units. The daily mean precipitation decreases by 0.093 units when we move from West to East on the same Latitude, whereas it increases by 0.035 units when we move further North on the same Longitude. 0.913 is the average measurement error . The total variation in the response(daily mean precipitation) is 52.6 units and beyond a distance of 31.44 (units) the data is no longer correlated.

3) Kriging

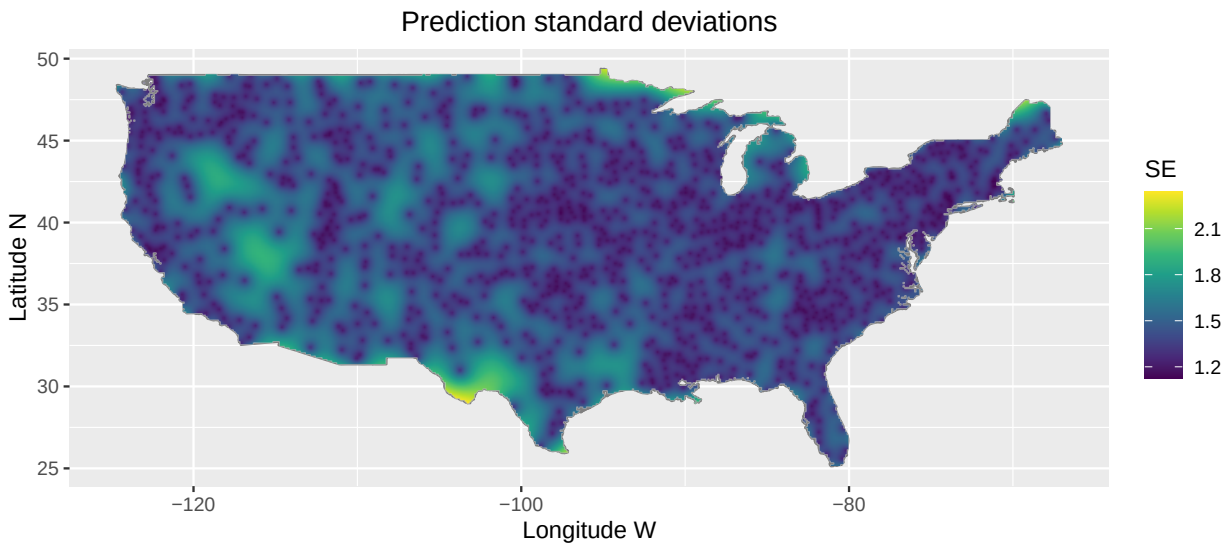
Plotting the observed mean precipitation:



Plotting the predicted mean precipitation from Model 1:



Plotting the standard deviations:



** Codes **

```
library(geoR)
library(maps)
library(ggplot2)
library(viridis)

#loading and cleaing data

load("C:/Users/sukan/Downloads/USHCNprcpSetup (2).RData")
PRCP_req <- PRCP[((years>=2000)&(years<=2009)),]

check_obs <- apply(PRCP_req,2,function(x){
  sum(!is.na(x))>=3000
})

PRCP_obs <- PRCP_req[,check_obs]

#mean precipitation
mean_dy <- apply(PRCP_obs,2,mean,na.rm=T)


lon    <- lon.lat[check_obs,1]
lat    <- lon.lat[check_obs,2]
Y <- mean_dy
X <- cbind(lon, lat)
s <- cbind(lon,lat)
n <- length(Y)

#initial values from variogram
init_sig2 <- 9
```

```
init_rho <- 7
init_tau2 <- 1.2

#5-fold cross validation
K      <- 5
set.seed(100)
fold   <- sample(1:K,n,replace=TRUE)
fold[1:5]

Yhat1 <- Yhat2 <- Yhat3 <- Yhat4<- sd1 <- sd2 <- sd3 <- sd4 <- rep(NA,n)

for(k in 1:K){
  train     <- fold != k
  test      <- fold == k

  fit_mle1  <- likfit(data=Y[train],trend= ~X[train,],coords=s[train,],
                      fix.nugget=FALSE,nugget=init_tau2,
                      cov.model="exponential",
                      ini = c(init_sig2, init_rho),messages=FALSE)

  fit_mle2  <- likfit(data=Y[train],trend= ~X[train,],coords=s[train,],
                      fix.nugget=TRUE,nugget=0,
                      cov.model="exponential",
                      ini = c(init_sig2, init_rho),messages=FALSE)
  fit_mle3  <- likfit(data=Y[train],trend= ~X[train,],coords=s[train,],
                      fix.nugget=FALSE,nugget=init_tau2,kappa=1.5,
                      cov.model="matern",
                      ini = c(init_sig2, init_rho),messages=FALSE)
  fit_mle4  <- likfit(data=Y[train],trend= ~X[train,],coords=s[train,],
                      fix.nugget=TRUE,nugget=0,kappa=1.2,
                      cov.model="matern",
                      ini = c(init_sig2, init_rho),messages=FALSE)
  pred1 <- krige.conv(data=Y[train],coords=s[train,], # Describe training data
                      locations=s[test,],              # Describe prediction sites
                      krige=krige.control(trend.d = ~X[train,],  # Covariates at s
                                          trend.l = ~X[test,],   # Covariates at s0
                                          cov.model="exponential",
                                          beta=fit_mle1$beta,
                                          cov.pars=fit_mle1$cov.pars,
                                          nugget=fit_mle1$nugget))
  pred2 <- krige.conv(data=Y[train],coords=s[train,],
                      locations=s[test,],
                      krige=krige.control(trend.d = ~X[train,],
                                          trend.l = ~X[test,],
                                          cov.model="exponential",
                                          beta=fit_mle2$beta,
                                          cov.pars=fit_mle2$cov.pars,
                                          nugget=fit_mle2$nugget))

  pred3 <- krige.conv(data=Y[train],coords=s[train,],
                      locations=s[test,],
                      krige=krige.control(trend.d = ~X[train,],
                                          trend.l = ~X[test,],
```

```
                                        cov.model="matern",
                                        beta=fit_mle3$beta,
                                        cov.pars=fit_mle3$cov.pars,
                                        nugget=fit_mle3$nugget,
                                        kappa=fit_mle3$kappa))

  pred4 <- krige.conv(data=Y[train],coords=s[train,],
                      locations=s[test,],
                      krige=krige.control(trend.d = ~X[train,],
                                        trend.l = ~X[test,],
                                        cov.model="matern",
                                        beta=fit_mle4$beta,
                                        cov.pars=fit_mle4$cov.pars,
                                        nugget=fit_mle4$nugget,
                                        kappa=fit_mle4$kappa))

  Yhat1[test] <- pred1$predict
  Yhat2[test] <- pred2$predict
  Yhat3[test] <- pred3$predict
  Yhat4[test] <- pred4$predict
  sd1[test]   <- sqrt(pred1$krige.var)
  sd2[test]   <- sqrt(pred2$krige.var)
  sd3[test]   <- sqrt(pred3$krige.var)
  sd4[test]   <- sqrt(pred4$krige.var)
}

MSE1   <- mean((Y-Yhat1)^2)
MSE2   <- mean((Y-Yhat2)^2)
MSE3   <- mean((Y-Yhat3)^2)
MSE4   <- mean((Y-Yhat4)^2)

MAD1   <- median(abs(Y-Yhat1))
MAD2   <- median(abs(Y-Yhat2))
MAD3   <- median(abs(Y-Yhat3))
MAD4   <- median(abs(Y-Yhat4))

SD1    <- mean(sd1)
SD2    <- mean(sd2)
SD3    <- mean(sd3)
SD4    <- mean(sd4)

COVER1 <- 100*mean(abs(Y-Yhat1)<1.96*sd1)
COVER2 <- 100*mean(abs(Y-Yhat2)<1.96*sd2)
COVER3 <- 100*mean(abs(Y-Yhat3)<1.96*sd3)
COVER4 <- 100*mean(abs(Y-Yhat4)<1.96*sd4)

COR1   <- cor(Y,Yhat1)
COR2   <- cor(Y,Yhat2)
COR3   <- cor(Y,Yhat3)
COR4   <- cor(Y,Yhat4)

#aic bic
fit_mle1  <- likfit(data=Y,trend= ~X,coords=s,
```

```
                    fix.nugget=FALSE,nugget=init_tau2,
                    cov.model="exponential",nospatial=TRUE,
                    ini = c(init_sig2, init_rho),messages=FALSE)
fit_mle2  <- likfit(data=Y,trend= ~X,coords=s,
                    fix.nugget=TRUE,nugget=0,
                    cov.model="exponential",nospatial=TRUE,
                    ini = c(init_sig2, init_rho),messages=FALSE)
fit_mle3  <- likfit(data=Y,trend= ~X,coords=s,
                    fix.nugget=FALSE,nugget=init_tau2,
                    cov.model="matern",fix.kappa = FALSE,nospatial=TRUE,
                    ini = c(init_sig2, init_rho),messages=FALSE)
fit_mle4  <- likfit(data=Y,trend= ~X,coords=s,
                    fix.nugget=TRUE,nugget=0,
                    cov.model="matern",fix.kappa = FALSE,nospatial=TRUE,
                    ini = c(init_sig2, init_rho),messages=FALSE)

mod1 <- c(fit_mle1$AIC, fit_mle1$BIC, MSE1, MAD1, SD1, COVER1, COR1)
mod2 <- c(fit_mle2$AIC, fit_mle2$BIC, MSE2, MAD2, SD2, COVER2, COR2)
mod3 <- c(fit_mle3$AIC, fit_mle3$BIC, MSE3, MAD3, SD3, COVER3, COR3)
mod4 <- c(fit_mle4$AIC, fit_mle4$BIC, MSE4, MAD4, SD4, COVER4, COR4)
model_att <- cbind(mod1,mod2,mod3,mod4)

#MLEs
Betas <- fit_mle1$beta

#kriging
s01 <- seq(-135,-65,0.1)
s02 <- seq(25,50,0.1)
s0  <- as.matrix(expand.grid(s01,s02))
dim(s0)
inusa <- map.where("usa",s0[,1],s0[,2])
s0    <- s0[!is.na(inusa),]
dim(s0)

X0     <- cbind(lon=s0[,1],lat=s0[,2])

pred <- krige.conv(data=Y,coords=s, # Describe training data
                   locations=s0,    # Describe prediction sites
                   krige=krige.control(trend.d = ~X,  # Covariates at s
                                       trend.l = ~X0, # Covariates at s0
                                       cov.model="exponential",
                                       beta=fit_mle1$beta,
                                       cov.pars=fit_mle1$cov.pars,
                                       nugget=fit_mle1$nugget
                                       ))

Yhat <- pred$predict         # Kriging predictions
se   <- sqrt(pred$krige.var) # Kriging prediction standard deviations

# Plot observed data
df1 <- data.frame(long=s[,1],lat=s[,2],ObservedMean=Y)
ggplot(df1, aes(long, lat)) +
  borders("state") +
```

```
  geom_point(aes(colour = ObservedMean)) +
  scale_colour_gradientn(colours = viridis(10)) +
  xlab("")+ylab("")+labs(title="Observed Mean Precipitation")+
  coord_fixed()

# Plot Kriging predictions
df2 <- data.frame(long=s0[,1],lat=s0[,2],PredMean=Yhat)
ggplot(df2, aes(long, lat)) +
  borders("state") +
  geom_raster(aes(fill = PredMean)) +
  scale_fill_gradientn(colours = viridis(10))+
  xlab("")+ylab("")+labs(title="Predicted Mean Precipitation")+
  coord_fixed()

# Plot Kriging standard deviations
df3 <- data.frame(long=s0[,1],lat=s0[,2],SE=se)
ggplot(df3, aes(long, lat)) +
  borders("state") +
  geom_raster(aes(fill = SE)) +
  scale_fill_gradientn(colours = viridis(10))+
  xlab("")+ylab("")+labs(title="Prediction standard deviations")+
  coord_fixed()
```