## ST540 Final Project

## Team Members: Xintong Jiang, Morgan Kuchenbrod, Leah Rohde, Yihang Xu

## Question 1

The data should be modeled as

where is the mean MAT for a given spatiotemporal coordinate and are independent errors. There are obviously other models for , but for this exam we will use splines. Let be a spline basis expansion of longitude, be a spline basis expansion of latitude and be a spline basis expansion of time. The mean MAT is modeled using all three sets of the basis functions and their interactions in a multiple linear regression:

Next we aim to specify the priors of these parameters and errors. The independent errors are assumed to be distributed as and is assigned a scaled inverse- prior with degree of freedom and scale parameter . Since , the dimension of can be rewritten as . The dimension of feature is .More precise, we assume that The prior is which has PDF

This is also known as the Bayesian LASSO prior. The remaining default parameters are fixed and same as package.

## Question 2

Define the the deviance as twice the negative log likelihood

Let be the posterior mean of the deviance. Denote as the posterior mean of . The effective number of parameters is

DIC can be written

We propose to use the Deviance information criteria (DIC) to select the number of basic functions. We will select the model with smallest DICs.

## Question 3

For simplicity, we always assume that . We investigate 5 models as . We use package to fit the model. The DICs of the model are

Hence, we choose the model. In such a case, the average MSE is

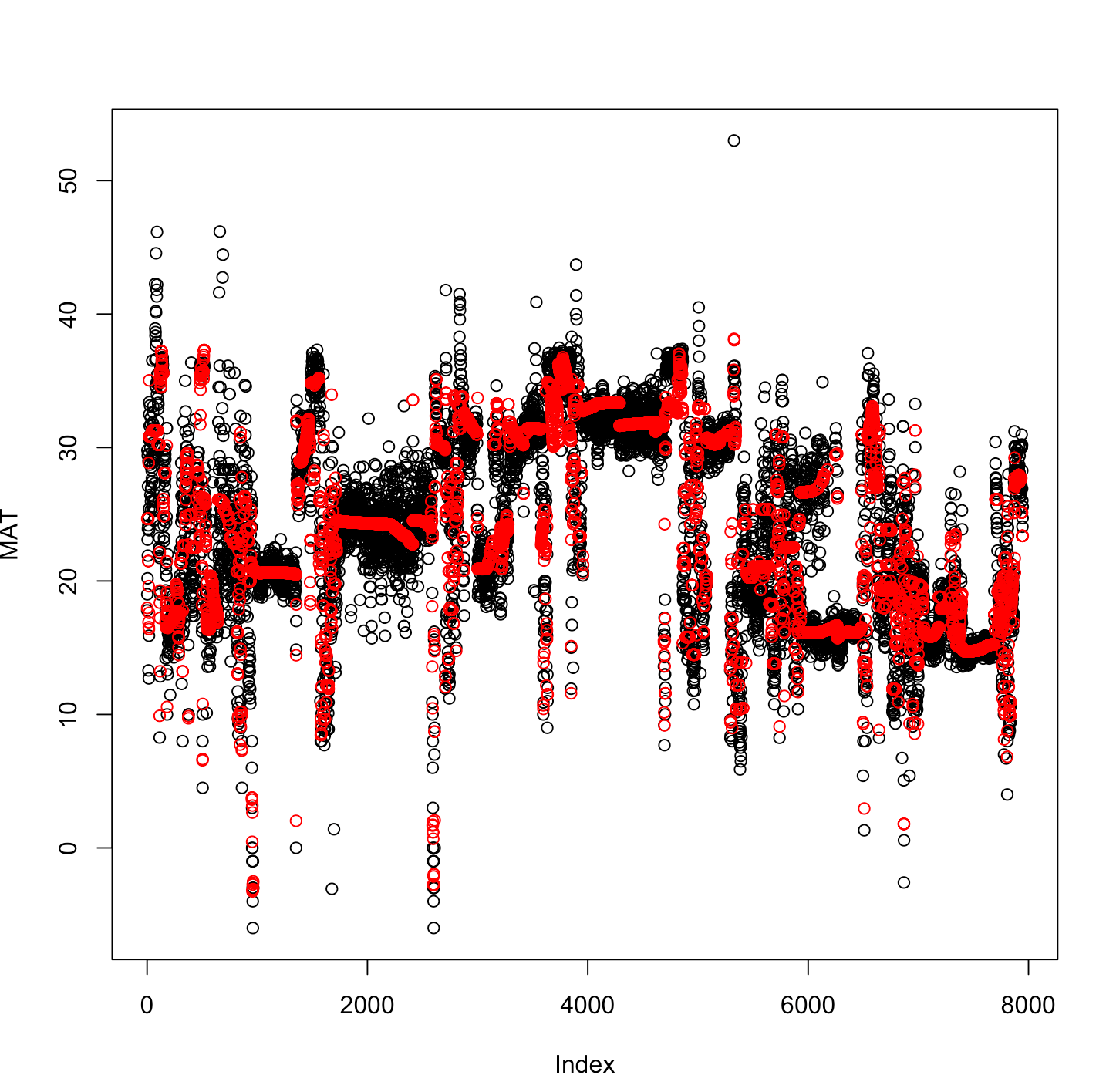
To further verify that the model fits well, we first plot the true MAT and the estimated MAT as follows. From Figure 1 below, we can find that BLR predicts the MAT well. The black points represent the true MAT and the red points represent the estimated MAT. 

Figure 1

## Question 4

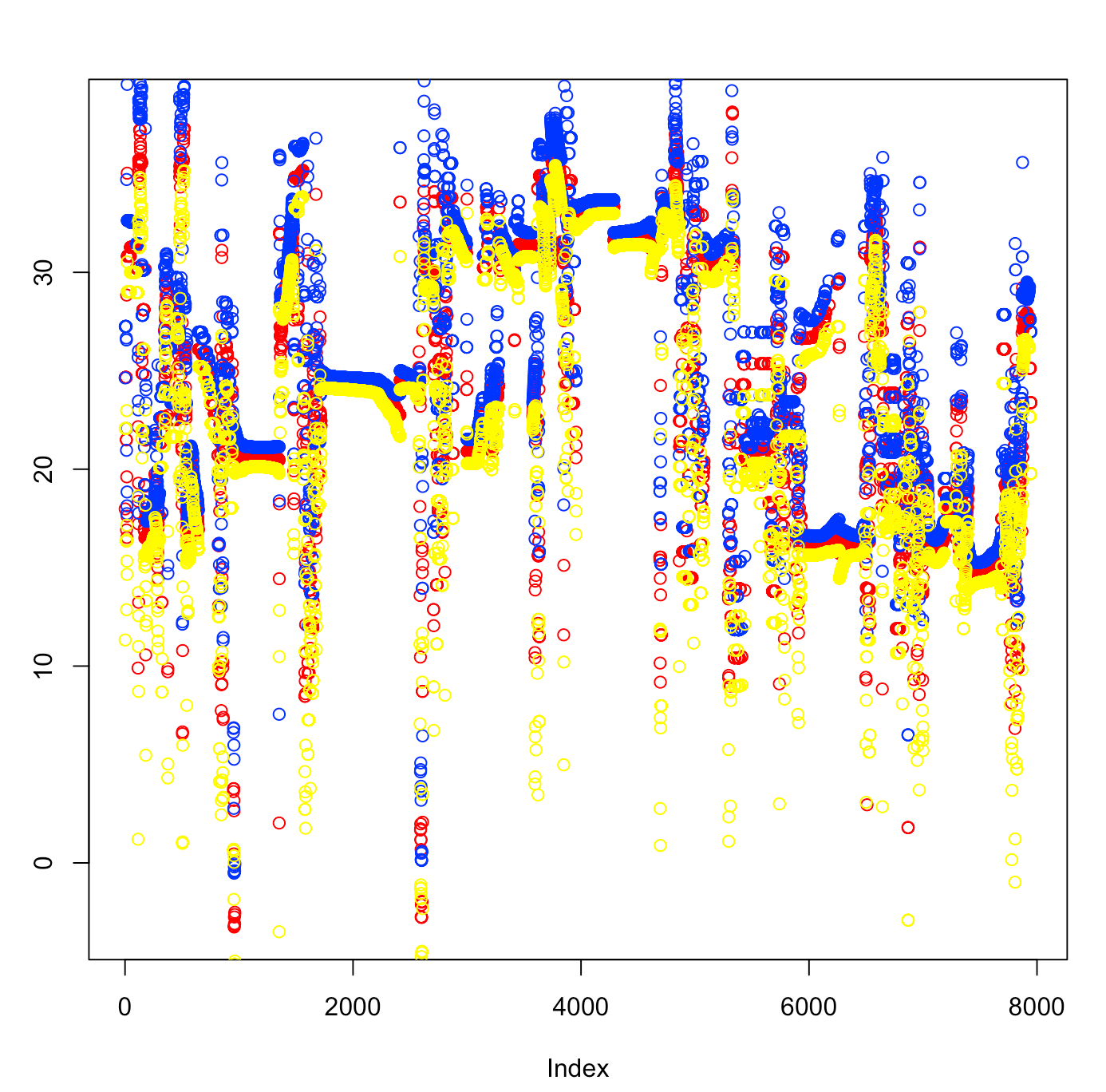
We use the posterior mean and the posterior standard variation to predict the value of MAT. We construct a 3D grid on the domain. The upper bound of the estimated MAT is set as and the lower bound of the estimated MAT is set as . The following Figure 2 shows the estimated MATs of observations. The red points represent the estimated MAT; The blue points represent the upper bound; The yellow points represent the lower bound. The remaining estimated MATs (3D grid) can be seen in Question 5. 

Figure 2

## Question 5

We successfully create 2 visualization tools that a user can input a time and see a map of estimated MAT or input a location and see a time series of estimated MAT.

Figure 3 showed an example of a 3D map when the time is age=102.63.

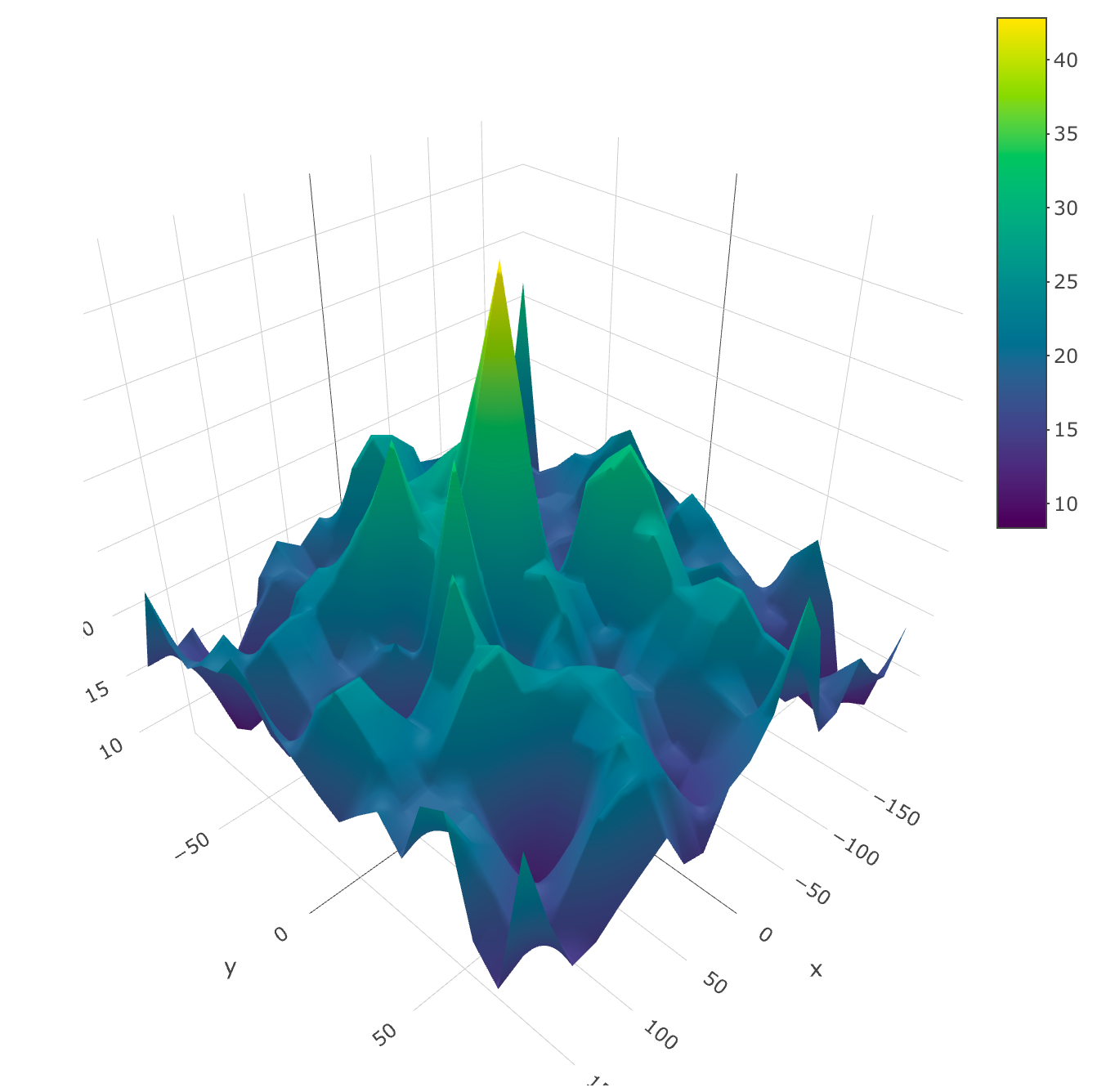


Figure 3. MATs (3D grid)

Figure 4 showed an example of plot for MAT corresponding to time series when the location is (lon=-9.473684, lat=-4.736842).

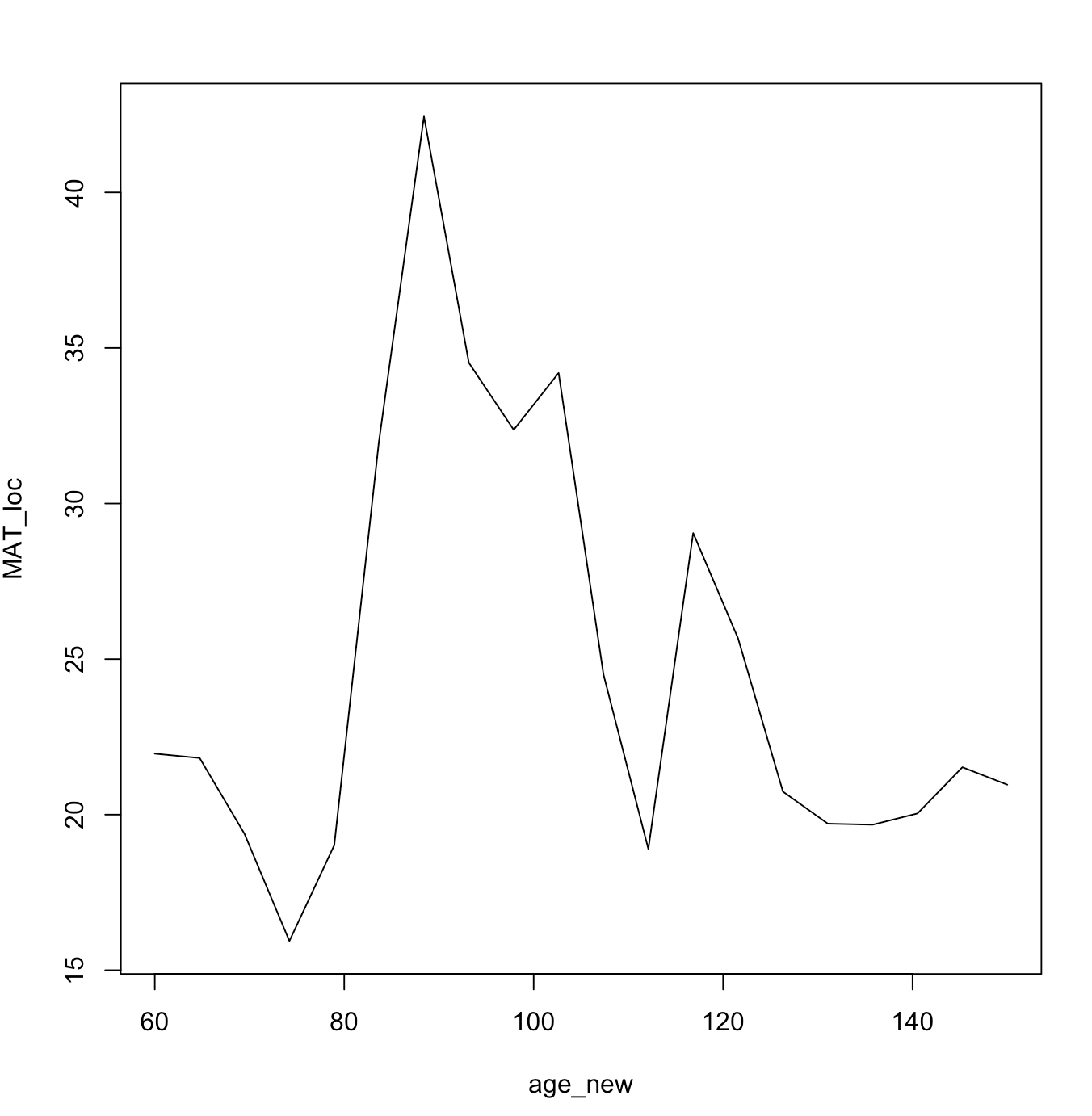


Figure 4

###Code Section

library(BGLR)

### Problem 1

### loading data data <- read.csv("C:/Users/yxu/Desktop/paleo\_dat.csv")

head(data)

data\_BLR = data[!is.na(data$Temperature.C),c(1,3,4,7)]

data\_BLR = data\_BLR[data\_BLR$Sample.Age>=60,]

head(data\_BLR)

### Fixing Model

bs\_fixed\_knots <- function(x, x\_min, x\_max, df) {

library(splines)

B = bs(x, df=df, Boundary.knots = c(x\_min,x\_max),

knots = seq(x\_min,x\_max,length = df -3)

)

return(B)

}

### Problem 2

### Use DIC as the strategy for selecting the number of basis functions

cal\_DIC <- function(knots) {

range1 = c(-180,180)

J = knots

lon = data\_BLR$Paleo.Lon

B1 = bs\_fixed\_knots(lon, range1[1], range1[2], J)

range2 = c(-90,90) K = knots lat = data\_BLR$Paleo.Lat

B2 = bs\_fixed\_knots(lat, range2[1], range2[2], K)

range3 = c(60,150)

L = knots age = data\_BLR$Sample.Age

B3 = bs\_fixed\_knots(age, range3[1], range3[2], L)

X = NULL

for (j in 1:J) {

for (k in 1:K) {

for (l in 1:L) {

X = cbind(X, B1[,j]\*B2[,k]\*B3[,l])

}

}

}

MAT = data\_BLR$Temperature.C

out=BLR(y=MAT,XL=X)

return(out$fit$DIC)

}

###### Problem 3 & 4

# Calculating DIC for several models

DIC\_vec = c(cal\_DIC(6), cal\_DIC(7), cal\_DIC(8), cal\_DIC(9), cal\_DIC(10))

DIC\_vec2 = c(cal\_DIC(11), cal\_DIC(12))

DIC\_vec3 = c(cal\_DIC(15), cal\_DIC(20))

# Further check whether the model with smallest DIC fit well by using plot.

knots = 20

range1 = c(-180,180)

J = knots

lon = data\_BLR$Paleo.Lon

B1 = bs\_fixed\_knots(lon, range1[1], range1[2], J)

range2 = c(-90,90)

K = knots

lat = data\_BLR$Paleo.Lat

B2 = bs\_fixed\_knots(lat, range2[1], range2[2], K)

range3 = c(60,150)

L = knots

age = data\_BLR$Sample.Age

B3 = bs\_fixed\_knots(age, range3[1], range3[2], L)

X = NULL

for (j in 1:J) {

for (k in 1:K) {

for (l in 1:L) {

X = cbind(X, B1[,j]\*B2[,k]\*B3[,l])

}

}

}

MAT = data\_BLR$Temperature.C

out=BLR(y=MAT,XL=X)

(mean((out$y - out$yHat)^2))

lon\_new = seq(-180,180,length = 20)

lat\_new = seq(-90,90,length = 20)

age\_new = seq(60,150,length = 20)

lon\_pre = NULL

lat\_pre = NULL

age\_pre = NULL

for (j in 1:20) {

for (k in 1:20) {

for (l in 1:20) {

lon\_pre = c(lon\_pre, lon\_new[j])

lat\_pre = c(lat\_pre,lat\_new[k])

age\_pre = c(age\_pre,age\_new[l])

}

}

}

MAT\_pre = rep(NA, 8000)

lon\_all = c(lon, lon\_pre)

lat\_all = c(lat, lat\_pre)

age\_all = c(age, age\_pre)

MAT\_all = c(MAT, MAT\_pre)

lengknots = 20

range1 = c(-180,180)

J = knots lon = data\_BLR$Paleo.Lon

B1 = bs\_fixed\_knots(lon\_all, range1[1], range1[2], J)

range2 = c(-90,90)

K = knots

lat = data\_BLR$Paleo.Lat

B2 = bs\_fixed\_knots(lat\_all, range2[1], range2[2], K)

range3 = c(60,150)

L = knots

age = data\_BLR$Sample.Age

B3 = bs\_fixed\_knots(age\_all, range3[1], range3[2], L)

X = NULL

for (j in 1:J) {

for (k in 1:K) {

for (l in 1:L) {

X = cbind(X, B1[,j]\*B2[,k]\*B3[,l])

}

}

}

out=BLR(y=MAT\_all,XL=X)

yhat = out$yHat

sdhat = out$SD.yHat

setwd("/Users/apple/Desktop/huge/TA/Baysian")

save(yhat, sdhat, file = "BLR.Rdata")

# Plot for problem 3

plot(MAT)

points(yhat[1:7947], col = 'red')

# Plot for problem 4

plot(yhat[1:7947], col="red")

points(yhat[1:7947]+3\*sdhat[1:7947], col = "blue")

points(yhat[1:7947]-3\*sdhat[1:7947], col = "yellow")

######## Problem 5

### plot with fixed time ###

time = age\_pre[10]

ind\_time = which(age\_pre == time)

lon\_time = lon\_pre[ind\_time]

lat\_time = lat\_pre[ind\_time]

MAT\_pre = yhat[7948:15947]

MAT\_time = matrix(MAT\_pre[ind\_time], ncol = 20,nrow = 20, byrow = T)

library(plot3D)

library(tidyverse)

library(plotly)

persp(lon\_new,lon\_new, MAT\_time)

plot\_ly(x = lon\_new, y = lat\_new, z = MAT\_time) %>% add\_surface()

### plot with fixed location (longitude and latitude) ###

location = c(lon\_new[10], lat\_new[10])

ind\_loc = which((lon\_pre == location[1]) & (lat\_pre == location[2]))

MAT\_pre = yhat[7948:15947]

MAT\_loc = MAT\_pre[ind\_loc]

plot(age\_new, MAT\_loc, type = "l")