ST 540 (602)

Midterm 2

Darlene Agle

4/11/2022

1.) **Model Definition:** Because the Enhanced Vegetation Index (EVI) exhibits strong seasonality, we expect our true EVI curve μ(t), to behave according to a non-linear pattern. I am prompted to use semiparametric regression to approximate this curve. It seems appropriate to add a Random Effect element to the model to compensate for correlation between samples taken in the same year, but as the samples were taken so inconsistently, at different times and with different frequencies each year, I was unable to construct an appropriate matrix to loop this in R. I settled on a model that does not include the random effect of year. This model operates under the assumption that EVI follows the same curve each year ($g_1(t_1)=g_2(t_2)=\ldots=g_n(t_n)$). After the model is built, I will evaluate the appropriateness of this assumption. My model: $Y_i|t_i$~Normal($g(t),\sigma^2$) where g(t) is the true EVI curve and is equal to $\mu+\sum_{j=1}^{J} B_j(t)\beta_j$ (spline basis representation). I fit a model with J=5 basis functions. I chose uninformative prior distributions for each of the hyperparameters:

μ_i~Normal(0,100) β_j~Normal($0,\sigma^2\tau^2$)   and σ^2,τ^2 ~InvGamma(0.1,0.1)

2.) **MCMC Convergence:** After 20000 iterations of Monte Carlo Sampling, I observed acceptable trace plots, but I observed too much autocorrelation between samples, even as the lag increased.

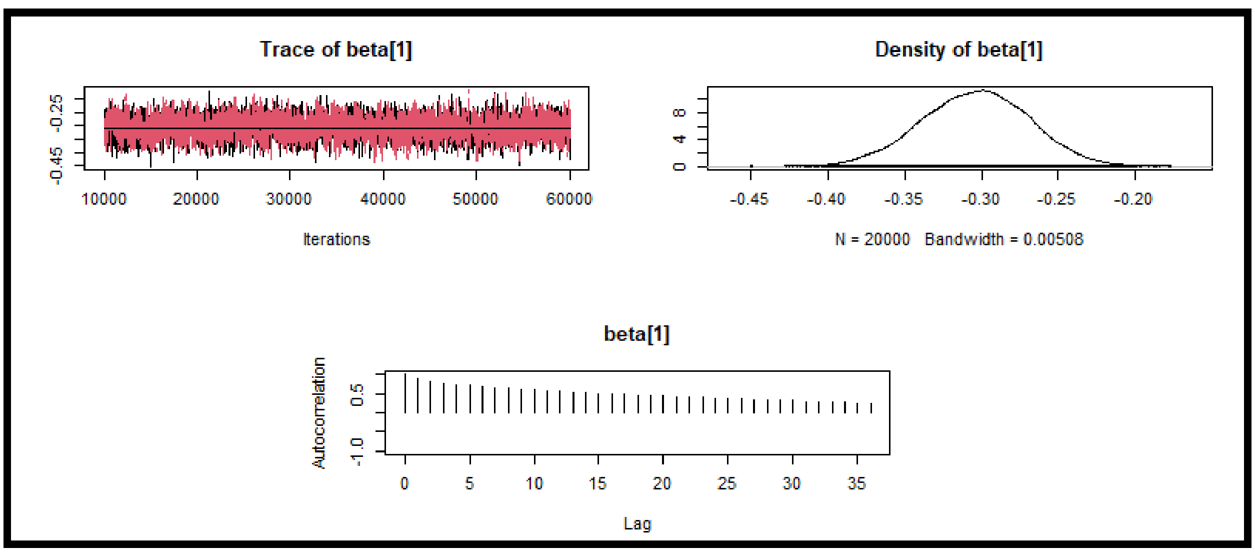

Figure1: Post burn-in trace plot, distribution, and autocorr. of 20000 MCMC samples for β1

I had to increase the number of iterations to 30000 to get an effective sample size of over 1000 for each parameter and strengthen my estimation of convergence.

3.) **Model comparisons**: Using Deviance Information Criteria, I compared this intricate model to one with only J=3 basis functions and one with J=10 basis functions.

DIC of Model with J=5 Basis Functions	**DIC of Model with J=3 Basis Functions**	**DIC of Model with J=10 Basis Functions**
Mean deviance: -1705	Mean deviance: -1166	Mean deviance: -2076
penalty 6.996	penalty 4.986	penalty 11.92
Penalized deviance: -1698	Penalized deviance: -1161	Penalized deviance: -2064

Table1: DIC of original model compared to a simpler and more complicated model

Though DIC favors the most complicated model here (a criticism of DIC), the curve under J=5 basis functions most closely resembles what we know about the rotation of the earth and has a DIC much lower than the J=3 model.

4.) **Model fit**: Under this model, assuming $\mu(t_1) = \mu(t2) = \ldots = \mu(t)$, the EVI curve behaves as follows each year:

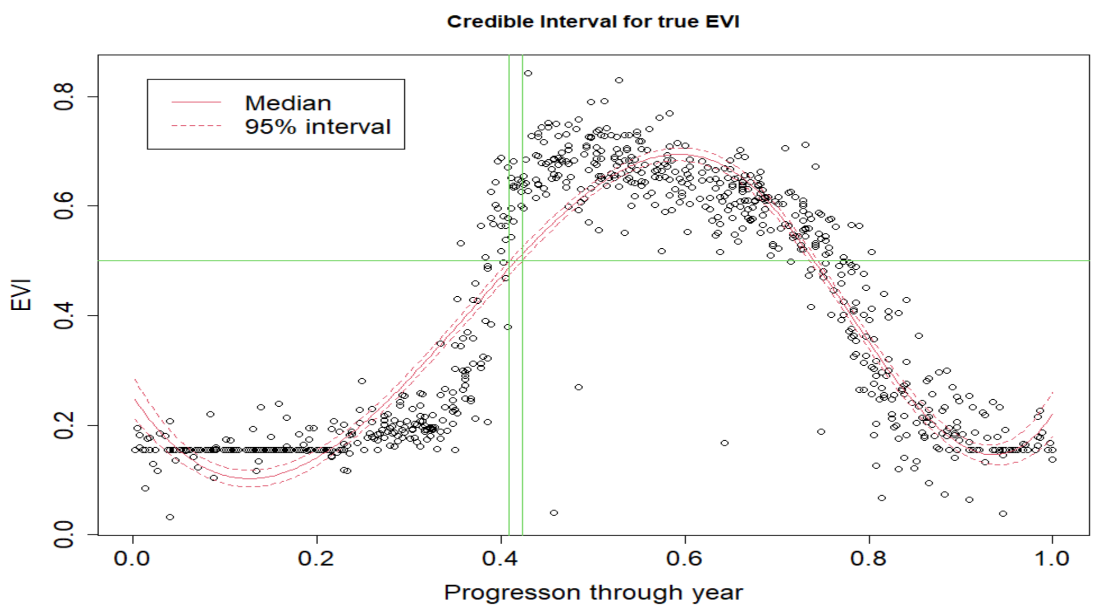

Figure 2: Median and 95% Credible Interval for $\mu(t)$, t scaled between 0 and 1, GUT credible interval marked in green.

According to this model, the 95% credible interval for GUT to occur is between day 149 and 165 (between May 29th and June 3rd).

5.) **GUT analysis**: Under my model, all posterior distributions for GUT are the same; However, since this does not get at the underlying question and is strictly due to my R limits, I decided to partition the data into 6 chronological groups and build the same model on each dataset.

	95% Credible Interval for GUT	
Time Period	Lower Bound	Upper Bound
1984-1994	Day 155	Day 166
1995-1999	Day 143	Day 159
2000-2004	Day 146	Day 157
2005-2009	Day 146	Day 157
2010-2014	Day 140	Day 154
2015-2019	Day 140	Day 153

Table 2: 95% Credible Interval for GUT based on 6 applications of my model to different sets of years

For each consecutive model, upper bounds for the credible interval of GUT decreased and the lower bounds generally decreased, suggesting that year is not a random effect but should be a factor in the model.

6.) **Time-trend analysis:** Based on the posterior distributions of the GUT for each of my 6 iterations of the semiparametric model, it appears there has been a change of GUT date across the years.

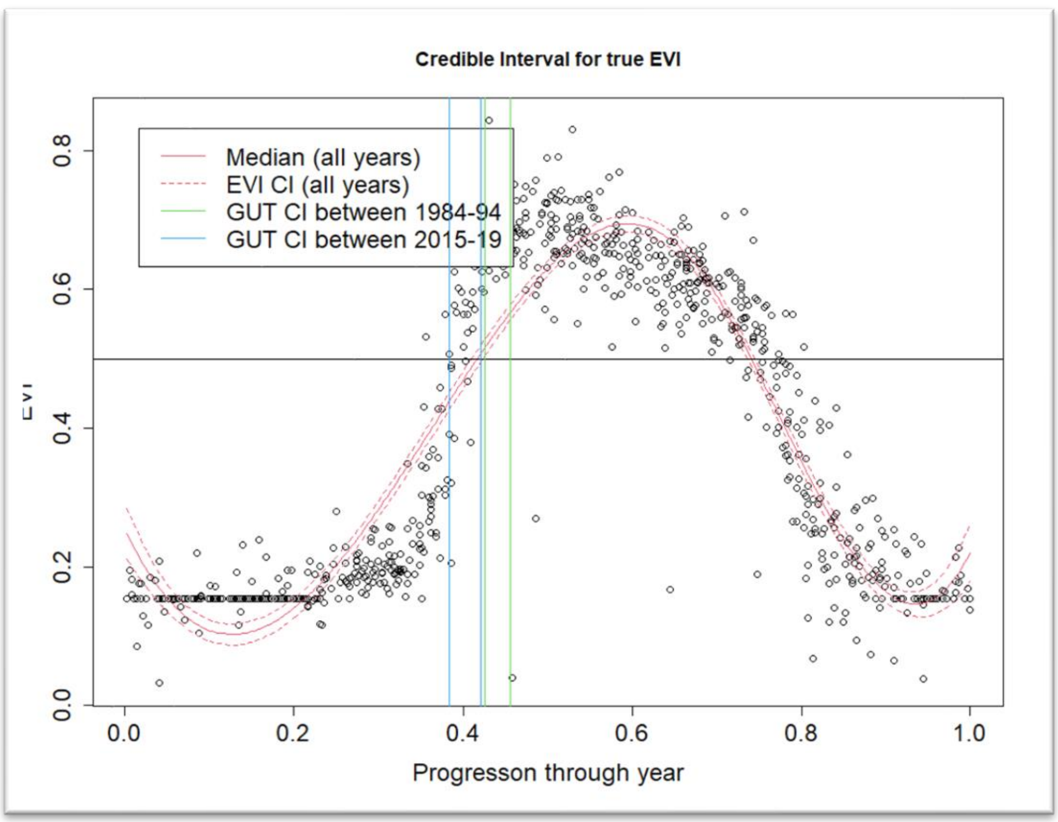

Figure 2: GUT CI's for the first group (green) and last group (blue)

The CI for GUT between the years 1984 and 1994 is outside of both the CI for GUT in the overall model and the CI for GUT for each of the other groups of years, suggesting that GUT used to occur earlier in the eighties and early nineties.

R Code:

```
library(MASS)
data<-read.csv('EVI_Data.csv',header=TRUE)
Y<- data$EVI
#Replaced with data$EVI[1:163],[164:233],[234:379],[380:525],[526:650],[651:802] to rerun model by group
X<- data$DOY
#Replaced with data$DOY[1:163],[164:233],[234:379],[380:525],[526:650],[651:802] to rerun model by group
X <- X/365

mat<-matrix(c(X,Y,Z),byrow=FALSE,ncol=3)
sorted.mat<-mat[order(mat[,1]),]
X<-sorted.mat[,1]
Y<-sorted.mat[,2]
Z<-sorted.mat[,3]
Z<-scale(Z)
n <- length(Y)
plot(X,Y,xlab="time",ylab="EVI",cex.lab=1.5,cex.axis=1.5)

library(splines)
J <- 5       # Number of basis functions (toggled between 3 and 10 for evaluation)
B <- bs(X,J)  

EVI_model <- "model{
   # Likelihood
   for(i in 1:n){
      Y[i]    ~ dnorm(mean[i],taue)
      mean[i] <-mu+inprod(B[i,],beta[])
   }
   # Prior
   mu   ~ dnorm(0,0.01)
   mu2~dnorm(0,0.01)
   taue ~ dgamma(0.1,0.1)
   for(j in 1:J){
    beta[j] ~ dnorm(0,taue*taub)
   }
   taub ~ dgamma(0.1,0.1)
  }"
library(rjags)
dat    <- list(Y=Y,n=n,B=B,J=J)
init   <- list(mu=mean(Y),beta=rep(0,J),taue=1/var(Y))
model  <- jags.model(textConnection(EVI_model),n.chains=2,
                     inits=init,data = dat,quiet=TRUE)
update(model, 10000, progress.bar="none")
samp   <- coda.samples(model, 
                       variable.names=c("mu","beta"), 
                       n.iter=30000, progress.bar="none")
samp2   <- coda.samples(model, 
                       variable.names=c("mean"), 
                       n.iter=30000, progress.bar="none")
sum <- summary(samp)
sum2<-summary(samp2)
plot(samp)
autocorr.plot(samp[1])
effectiveSize(samp)

q <- sum2$quantiles
plot(X,Y,xlab="Progresson through year",ylab="EVI",main="Credible Interval for true EVI",
     cex.lab=1.5,cex.axis=1.5)

lines(X,q[,1],col=2,lty=2)
lines(X,q[,3],col=2,lty=1)
lines(X,q[,5],col=2,lty=2)

legend("topleft",c("Median (all years)","EVI CI (all years)"),
       lty=c(1,2,1,1),col=c(2,2,3,4),bg=gray(1),inset=0.05,cex=1.5)
abline(h=.5,col=1)
results<-matrix(c(X,q[,1],q[,3],q[,5]),byrow=TRUE, nrow=4)

# Compute DIC
DIC    <- dic.samples(model,n.iter=30000, progress.bar="none")
DIC
```