ST440/540 – Applied Bayesian Analysis:
Midterm exam 2
Tilekbek Zhoroev

1 Model definition

Based on the data, we assume that Y_i are normally distributed regularly distributed on

$$Y_i \sim \mathcal{N}(\mu(t), \tau)$$

where τ is precision. We can employ a non-parametric statistical model because we have supplied few sample sizes for several years. The data matches well when we apply $B-$spline, however it's difficult to pinpoint the "green-up time" (GUT) each year. Then, we look at the logistic curve model. The general logistic function

$$f(t) = \frac{L}{1 + e^{-a(t-t_0)}}$$

, where L represents total capacity, a represents growth rate, and t_0 represents time when $f(t)$ equals $\frac{L}{2}$. By motivate this function, we define $\mu(t)$ as the difference between two logistic functions with $L = 1$,(since maximum value of $\mu(t)$ is 1) which first one captures an increasing pattern and the second one catches a drop pattern. So, we have

$$\mu(t) = \beta_0 + \frac{1}{1 + e^{-\beta_2(t-\beta_1)}} - \frac{1}{1 + e^{-\beta_4(t-\beta_3)}}.$$

The priors of the parameters are,

$$\begin{aligned}
&\beta_1 \sim \mathcal{N}(\mu_{\beta_1}, \tau)_{\chi(0,\infty)}, && \mu_{\beta_1} \sim \mathcal{U}(0, 183)\\
&\beta_2 \sim \mathcal{N}(\mu_{\beta_2}, \tau)_{\chi(0,\infty)}, && \mu_{\beta_2} \sim \mathcal{U}(0.1, 20)\\
&\beta_3 \sim \mathcal{N}(\mu_{\beta_3}, \tau)_{\chi(0,\infty)}, && \mu_{\beta_3} \sim \mathcal{U}(182, 360)\\
&\beta_4 \sim \mathcal{N}(\mu_{\beta_4}, \tau)_{\chi(0,\infty)}, && \mu_{\beta_4} \sim \mathcal{U}(0.1, 20)\\
&\beta_0 \sim \mathcal{N}(0, 0.1)_{\chi(0,\infty)}, && \tau \sim \Gamma(0.1, 0.1)
\end{aligned}$$

Because of the meanings in the logistic curve and physical constraints, we chose these priors.

2 MCMC convergence

We used MCMC to draw samples from posterior distributions with aggregated data. The chain plots in Figure 1 verify that the MCMC converged and it's enough to explore the posterior distributions. The effective sample sizes are fairly large (over 1125), and the Gelman-Rubin statistics are close to 1 and the magnitude of Geweke statistics is less than 2, as shown in Table 1. Therefore, the chains have clearly converged.

parameters	Effective size	Gelman diagnostics		Geweke diagnostics
		Point Est.	Upper C.I.	
β_0	1818.457	1	1.01	1.7896
β_1	1912.192	1	1.01	-0.4989
β_2	1366.232	1	1.01	1.1287
β_3	1127.135	1	1.01	-1.4490
β_4	1613.982	1	1.01	-0.8184

Table 1: MCMC convergence diagnostics.

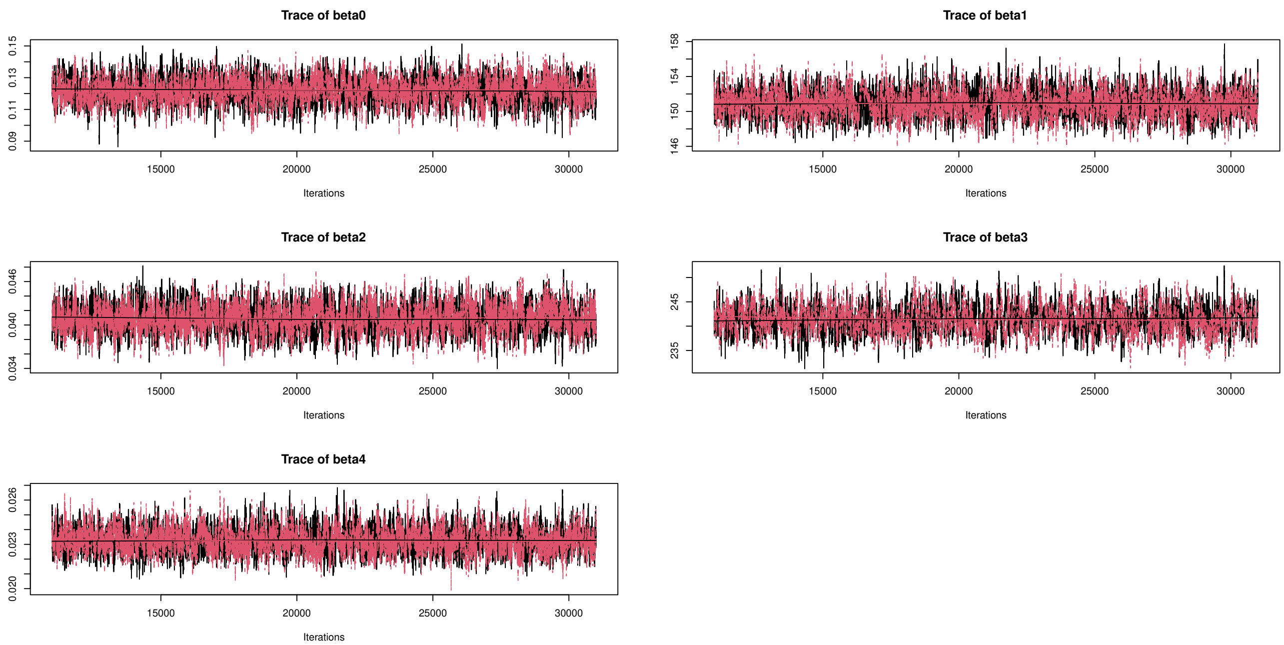

Figure 1: MCMC chain plot of the parameters.

3 Model comparisons

We compared three models with $\beta_0 = 0$, $\beta_0 \sim \mathcal{N}(0, 0.1)_{\chi(0,\infty)}$, and $\beta_0 \sim$ LogNormal$(0, 0.1)$. Both WAIC and DIC show strong support for Model 2, as shown in Table 2. It's worth mentioning that because Model 2 is the simplest, it has the lowest penalty score.

	DIC mean	DIC penalty	WAIC
Model 1	-1582	5.66	-1448.76
Model 2	-1797	6.613	-1787.31
Model 3	-1715	7.24	-1615.36

Table 2: Model comparison values across the three models

4 Model fit

We fit the data to each year after selecting the model in the previous scenario. We showed model fits for some year's data in Figure 2. We can see from the graph that 95% credible intervals cover the majority of the data, and the level of uncertainty varies depending on the number of data and time.

5 GUT analysis

From logistic curve we know that parameter β_1 give us the GUT. So, after fitting data to the model we obtain the posterior samples of $beta_1$ and it's given in Table 3.

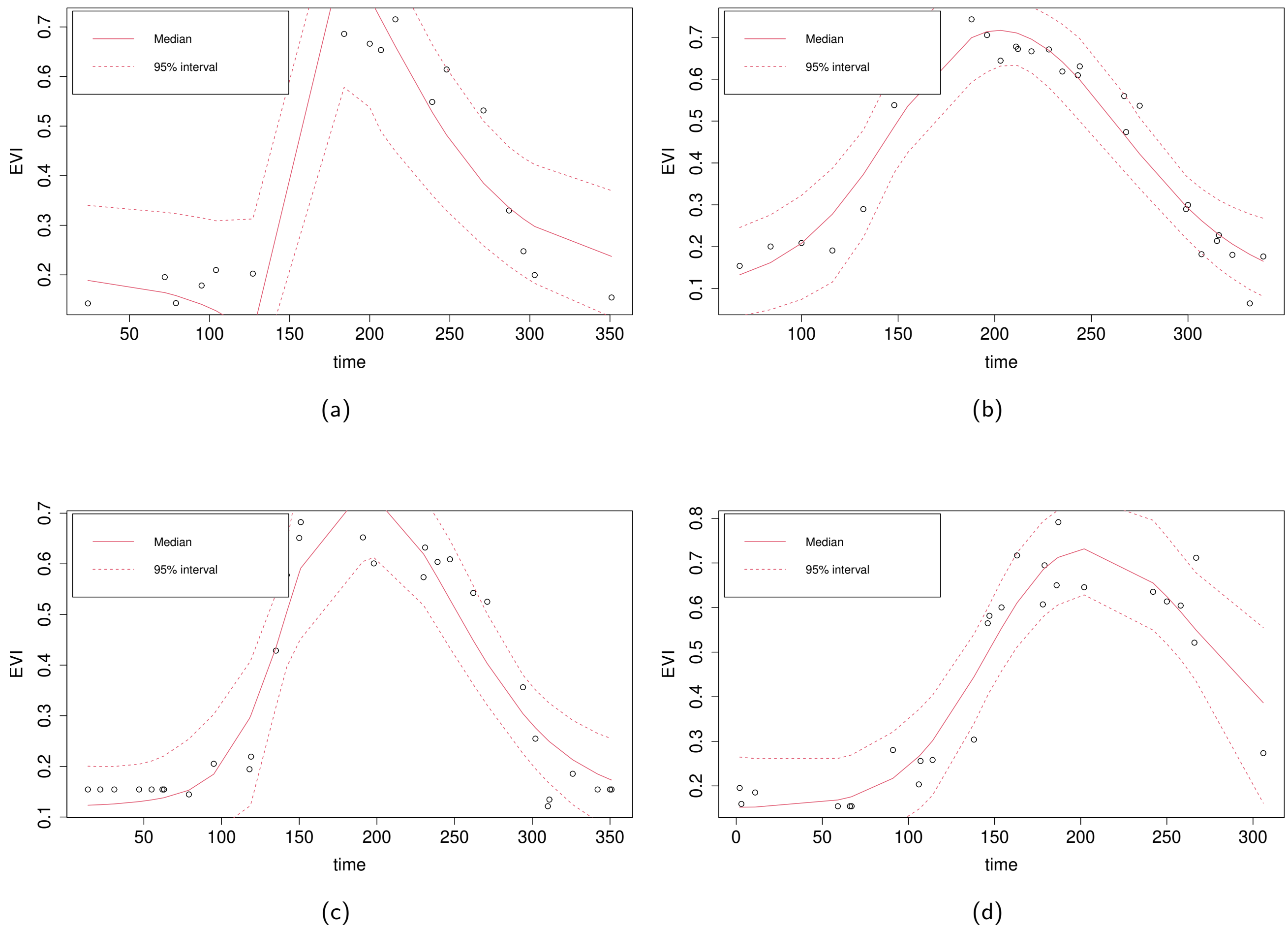

Figure 2: Plot of the $\mu(t)$, 95% credible interval of $\mu(t)$, and EVI data in (a) 1989; (b) 1999; (c) 2009; (d) 2019.

years	**1984**	**1985**	**1986**	**1987**	**1988**	**1989**	**1990**	**1991**	**1992**	**1993**	**1994**	**1995**
2.5%	12	45	30	140	135	128	7	117	74	65	143	128
Median	129	168	206	171	218	150	150	137	205	114	183	148
97.5%	240	241	247	243	336	182	232	257	247	185	237	228
years	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**
2.5%	6	48	116	130	132	133	134	138	124	136	134	128
Median	92	176	190	151	155	145	155	162	145	154	146	146
97.5%	207	245	248	181	178	161	181	181	187	179	176	171
year	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
2.5%	132	129	116	136	79	127	133	128	122	131	132	122
Median	145	143	128	164	139	151	151	145	151	147	143	149
97.5%	162	176	147	242	238	192	183	180	191	166	164	180

Table 3: Summary of posterior distribution of GUT; median and 95% credible interval of GUT for each year

6 Time-trend analysis

We can see that the posterior median and credible interval fluctuate by year using a summary of the posterior distribution of GUT in Table 3; for example, the median in 1996 is 92, whereas the median in 1988 is 218. Furthermore, by comparing the posterior distributions of 1987 and 2010, we can see that the credible intervals only include a few time steps.

Appendix

```
EVI_Data <- read_csv("EVI_Data.csv")
Y = EVI_Data$EVI; time = EVI_Data$DOY; years = EVI_Data$Year;
uniq_years = unique(years); num_years = length(uniq_years); N = length(Y)
ind = order(time)
time = time[ind]
Y = Y[ind]
evi_model <- "model{

   # Likelihood
   for(i in 1:n){
      Y[i] ~ dnorm(mean[i],taum)
      mean[i] <- beta0 + 1/(1+exp(-beta2*(t[i]-beta1))) - 1/(1+exp(-beta4*(t[i]-beta3)))
   }

   # Prior
   beta1  ~ dnorm(mu1,tau1)T(0,)
   beta2 ~ dnorm(mu2,tau2)T(0,)
   beta3 ~ dnorm(mu3,tau3)T(0,)
   beta4 ~ dnorm(mu4,tau4)T(0,)
   mu1 ~ dunif(0, 183)
   mu3 ~ dunif(182, 360)
   mu2 ~ dunif(0.1, 20)
   mu4 ~ dunif(0.1, 20)
   beta0 ~dnorm(0,0.01)T(0,)
   taum ~ dgamma(0.1,0.1)
   tau1 ~ dgamma(0.1,0.1)
   tau2 ~ dgamma(0.1,0.1)
   tau3 ~ dgamma(0.1,0.1)
   tau4 ~ dgamma(0.1,0.1)
  }"
dat    <- list(Y=Y,n=N,t =time)
init   <- list(mu1=96, mu3 = 271, mu2 = 5, mu4 = 5)
model  <- jags.model(textConnection(evi_model), inits=init,data = dat,n.chains=2,quiet=TRUE)
update(model, 10000, progress.bar="none")
samp   <- coda.samples(model,
             variable.names=c("beta0","beta1","beta2","beta3","beta4"),
             n.iter=20000, progress.bar="none")
plot(samp,density = False)
effectiveSize(samp)
gelman.diag(samp)
```

```
geweke.diag(samp[[1]])

# Model 1
evi_model <- "model{

   # Likelihood
   for(i in 1:n){
      Y[i] ~ dnorm(mean[i],taum)
      mean[i] <- 1/(1+exp(-beta2*(t[i]-beta1))) - 1/(1+exp(-beta4*(t[i]-beta3)))
   }

   # Prior
   beta1  ~ dnorm(mu1,tau1)T(0,)
   beta2 ~ dnorm(mu2,tau2)T(0,)
   beta3 ~ dnorm(mu3,tau3)T(0,)
   beta4 ~ dnorm(mu4,tau4)T(0,)
   mu1 ~ dunif(0, 183)
   mu3 ~ dunif(182, 360)
   mu2 ~ dunif(0.1, 20)
   mu4 ~ dunif(0.1, 20)
   taum ~ dgamma(0.1,0.1)
   tau1 ~ dgamma(0.1,0.1)
   tau2 ~ dgamma(0.1,0.1)
   tau3 ~ dgamma(0.1,0.1)
   tau4 ~ dgamma(0.1,0.1)
   # WAIC calculations
   for(i in 1:n){
     like[i] = dnorm(Y[i],mean[i],taum)
   }
  }"
   dat    <- list(Y=Y,n=N,t =time)
   init   <- list(mu1=96, mu3 = 271, mu2 = 5, mu4 = 5)
   model  <- jags.model(textConnection(evi_model),
                      inits=init,data = dat,n.chains=2,quiet=TRUE)

   update(model, 10000, progress.bar="none")

   samp   <- coda.samples(model,
             variable.names=c("like"),
             n.iter=20000, progress.bar="none")
   # Compute DIC
DIC_logit  <- dic.samples(model,n.iter=20000,progress.bar="none")

# Compute WAIC
like       <- rbind(samp[[1]],samp[[2]]) # Combine the two chains
fbar       <- colMeans(like)
Pw         <- sum(apply(log(like),2,var))
WAIC_logit <- -2*sum(log(fbar))+2*Pw
DIC_logit
WAIC_logit
```

```r
# model 2

evi_model <- "model{

   # Likelihood
   for(i in 1:n){
      Y[i] ~ dnorm(mean[i],taum)
      mean[i] <- beta0 + 1/(1+exp(-beta2*(t[i]-beta1))) - 1/(1+exp(-beta4*(t[i]-beta3)))
   }

   # Prior
   beta1  ~ dnorm(mu1,tau1)T(0,)
   beta2 ~ dnorm(mu2,tau2)T(0,)
   beta3 ~ dnorm(mu3,tau3)T(0,)
   beta4 ~ dnorm(mu4,tau4)T(0,)
   mu1 ~ dunif(0, 183)
   mu3 ~ dunif(182, 360)
   mu2 ~ dunif(0.1, 20)
   mu4 ~ dunif(0.1, 20)
   beta0 ~dnorm(0,0.01)T(0,)
   taum ~ dgamma(0.1,0.1)
   tau1 ~ dgamma(0.1,0.1)
   tau2 ~ dgamma(0.1,0.1)
   tau3 ~ dgamma(0.1,0.1)
   tau4 ~ dgamma(0.1,0.1)
   # WAIC calculations
   for(i in 1:n){
     like[i] = dnorm(Y[i],mean[i],taum)
   }
  }"
   dat    <- list(Y=Y,n=N,t =time)
   init   <- list(mu1=96, mu3 = 271, mu2 = 5, mu4 = 5)
   model  <- jags.model(textConnection(evi_model),
                      inits=init,data = dat,n.chains=2,quiet=TRUE)

   update(model, 10000, progress.bar="none")

   samp   <- coda.samples(model,
             variable.names=c("like"),
             n.iter=20000, progress.bar="none")
   # Compute DIC
DIC_logit  <- dic.samples(model,n.iter=20000,progress.bar="none")

# Compute WAIC
like        <- rbind(samp[[1]],samp[[2]]) # Combine the two chains
fbar        <- colMeans(like)
Pw          <- sum(apply(log(like),2,var))
WAIC_logit <- -2*sum(log(fbar))+2*Pw
DIC_logit
```

```r
WAIC_logit


# Model 3

evi_model <- "model{

   # Likelihood
   for(i in 1:n){
      Y[i] ~ dnorm(mean[i],taum)
      mean[i] <-beta0 + 1/(1+exp(-beta2*(t[i]-beta1))) - 1/(1+exp(-beta4*(t[i]-beta3)))
   }

   # Prior
   beta0 ~ dlnorm(0,0.01)
   beta1  ~ dnorm(mu1,tau1)T(0,)
   beta2 ~ dnorm(mu2,tau2)T(0,)
   beta3 ~ dnorm(mu3,tau3)T(0,)
   beta4 ~ dnorm(mu4,tau4)T(0,)
   mu1 ~ dunif(0, 183)
   mu3 ~ dunif(182, 360)
   mu2 ~ dunif(0.1, 20)
   mu4 ~ dunif(0.1, 20)
   taum ~ dgamma(0.1,0.1)
   tau1 ~ dgamma(0.1,0.1)
   tau2 ~ dgamma(0.1,0.1)
   tau3 ~ dgamma(0.1,0.1)
   tau4 ~ dgamma(0.1,0.1)
   # WAIC calculations
   for(i in 1:n){
     like[i]  = dnorm(Y[i],mean[i],taum)
   }
  }"
   library(rjags)
   dat    <- list(Y=Y,n=N,t =time)
   init   <- list(mu1=96, mu3 = 271, mu2 = 5, mu4 = 5)
   model  <- jags.model(textConnection(evi_model),
                        inits=init,data = dat,n.chains=2,quiet=TRUE)

   update(model, 10000, progress.bar="none")

   samp1   <- coda.samples(model,
             variable.names=c("like"),
             n.iter=20000, progress.bar="none")
   # Compute DIC
DIC_logit  <- dic.samples(model,n.iter=20000,progress.bar="none")

# Compute WAIC
like       <- rbind(samp1[[1]],samp1[[2]]) # Combine the two chains
fbar       <- colMeans(like)
```

```
Pw         <- sum(apply(log(like),2,var))
WAIC_logit <- -2*sum(log(fbar))+2*Pw
DIC_logit
WAIC_logit


# Model fits

# change year

Y = EVI_Data[EVI_Data$Year == 2019,]$EVI; time = EVI_Data[EVI_Data$Year == 2019,]$DOY;
N = length(Y)
evi_model <- "model{

   # Likelihood
   for(i in 1:n){
      Y[i] ~ dnorm(mean[i],taum)
      mean[i] <- beta0 + 1/(1+exp(-beta2*(t[i]-beta1))) - 1/(1+exp(-beta4*(t[i]-beta3)))
   }

   # Prior
   beta1  ~ dnorm(mu1,tau1)T(0,250)
   beta2 ~ dnorm(mu2,tau2)T(0,)
   beta3 ~ dnorm(mu3,tau3)T(0,)
   beta4 ~ dnorm(mu4,tau4)T(0,)
   mu1 ~ dunif(0, 183)
   mu3 ~ dunif(182, 360)
   mu2 ~ dunif(0.1, 20)
   mu4 ~ dunif(0.1, 20)
   beta0 ~dnorm(0,0.01)T(0,)
   taum ~ dgamma(0.1,0.1)
   tau1 ~ dgamma(0.1,0.1)
   tau2 ~ dgamma(0.1,0.1)
   tau3 ~ dgamma(0.1,0.1)
   tau4 ~ dgamma(0.1,0.1)
  }"
   library(rjags)
   dat    <- list(Y=Y,n=N,t =time)
   init   <- list(mu1=96, mu3 = 271, mu2 = 5, mu4 = 5)
   model  <- jags.model(textConnection(evi_model),
                     inits=init,data = dat,n.chains=2,quiet=TRUE)

   update(model, 10000, progress.bar="none")

   samp   <- coda.samples(model,
             variable.names=c("mean"),
             n.iter=20000, progress.bar="none")
   sum <- summary(samp)
   q <- sum$quantiles
```

```
  plot(time,Y,xlab="time",ylab="EVI",
       cex.lab=1.5,cex.axis=1.5)

  lines(time,q[,1],col=2,lty=2) # 0.025 quantile (lower bound)
  lines(time,q[,3],col=2,lty=1) # 0.500 quantile (median)
  lines(time,q[,5],col=2,lty=2) # 0.975 quantile (upper bound)

  legend("topleft",c("Median","95% interval"),
         lty=1:2,col=2,bg=gray(1),inset=0.01,cex=1)

# Posterior distribution summary of GUT


Y = EVI_Data[EVI_Data$Year == 1987,]$EVI; time = EVI_Data[EVI_Data$Year ==1987,]$DOY;
N = length(Y)
evi_model <- "model{

  # Likelihood
  for(i in 1:n){
    Y[i] ~ dnorm(mean[i],taum)
    mean[i] <- beta0 + 1/(1+exp(-beta2*(t[i]-beta1))) - 1/(1+exp(-beta4*(t[i]-beta3)))
  }

  # Prior
  beta1  ~ dnorm(mu1,tau1)T(0,250)
  beta2 ~ dnorm(mu2,tau2)T(0,)
  beta3 ~ dnorm(mu3,tau3)T(0,)
  beta4 ~ dnorm(mu4,tau4)T(0,)
  mu1 ~ dunif(0, 183)
  mu3 ~ dunif(182, 360)
  mu2 ~ dunif(0.1, 20)
  mu4 ~ dunif(0.1, 20)
  beta0 ~dnorm(0,0.01)T(0,)
  taum ~ dgamma(0.1,0.1)
  tau1 ~ dgamma(0.1,0.1)
  tau2 ~ dgamma(0.1,0.1)
  tau3 ~ dgamma(0.1,0.1)
  tau4 ~ dgamma(0.1,0.1)
 }"
  library(rjags)
  dat    <- list(Y=Y,n=N,t =time)
  init   <- list(mu1=96, mu3 = 271, mu2 = 5, mu4 = 5)
  model  <- jags.model(textConnection(evi_model),
                       inits=init,data = dat,n.chains=2,quiet=TRUE)

  update(model, 10000, progress.bar="none")

  samp   <- coda.samples(model,
            variable.names=c("beta1"),
            n.iter=20000, progress.bar="none")
```

```r
sum <- summary(samp)
q <- sum$quantiles
q
```