

Guanqian-Wang-STAT6570-final-project

Guanqian Wang

Introduction

1.1 General Background and Scientific Question

Wild salmon are very treasured resources. A better understanding on its reproducing pattern could provide valuable information. In Pacific Northwest, salmon spawns are born in sites with river(non-ocean) water. Later, they will leave to ocean. But, for reproducing, salmon will swim back to their spawn sites and lay eggs. After those eggs hatched, those young salmon will repeat their parents and ancestors' life pattern again.

In this report, it will study the ratio of the number of recruitment (new offspring of salmon) over number of spawns (adult salmon going to back to site and reproduce). More specifically, the report will investigate what factors may influence this ratio and how those factors influenced this ratio. This ratio is interesting, because it reflects the efficiency/productivity of salmon's reproducing. One may interpret it as how many offspring per spawn has. The report will only study the ratio for pink salmon species.

1.2 Data Introduction

There are two datasets for this report.

The first dataset is "pinkdata", which includes observations of recruitment (named R in dataset), spawn (named S in data set) from 43 spawn site (named stock in the dataset) and 14 regions, over year 1950 to year 1996. Each observation also included information of latitude, longitude and along shore distance.

Another dataset is "stockinfo", which includes each 43 stocks' region, latitude, longitude and along shore distance. This dataset also included stocks of non-pink salmon species; this report will not use those observations. Most of the information in this dataset is consistent with the dataset "pinkdata", except a few observations. There's also some rounding difference of latitude and longitude from two dataset.

1.3 Missing Data and Inconsistent Observations

For the dataset "pinkdata", it appears there's some missing data of longitude, latitude and along shore distance for two stocks: Puyallup and Snohomish. But, the dataset "stockinfo" has these data. So, those missing values should not cause any trouble for analysis.

However, there're some inconsistency for those two stocks between two datasets. For the variable "Region", the dataset "pinkdata" shows Snohomish and part of Puyallup are in Puget Sound region while the "stockinfo" dataset indicates they are all in Washington region. It is unclear why this is happened.

Yet, this may not be a sever problem. "Region" is one of geographical location information. There are many consistent geographical information variables like latitude, longitude and along shore distance in the data set. Thus, this report simply chooses to not use the variable "Region".

1.4 Some Data Manipulation

Some new variables are introduced in the dataset.

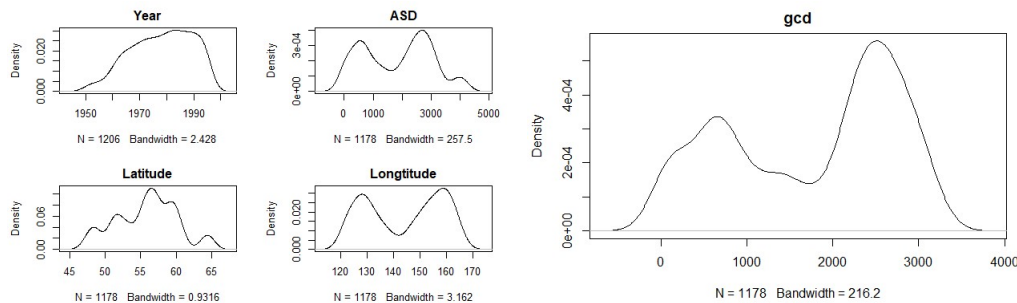
The first variable introduced is Y, which is $\log(\text{Recruitment}/\text{Spawn})$. The ratio Recruitment/Spawn is the report's variable of interest. The reason to take a log of it is to bring the scale down and make its distribution closer to the bell shape/normal distribution.

Another variable introduced is “gcd”, the Great Circle Distance. This variable is computed through using latitude and longitude. It is great circle distance between each observation and the most south observation. The computation of this variable was already specified in resource file. In a nutshell, this distance is the shortest distance between two points on surface of earth. This variable should be very similar to variable “along shore distance”. According to the final project instruction, “along shore distance” also has value 0 for the most south observation, and the rest observations' values are becoming larger from south to north.

The last variable introduce is “Yr_type”, it is an indicator variable to differentiate odd year and even year. It will be 1 if it's even year and will be 0 if it's odd year. The reason to introduce this variable is because according to the final project instruction, salmons returning at odd year and even year are different species. It is reasonable to consider that different species may have different reproducing productivity.

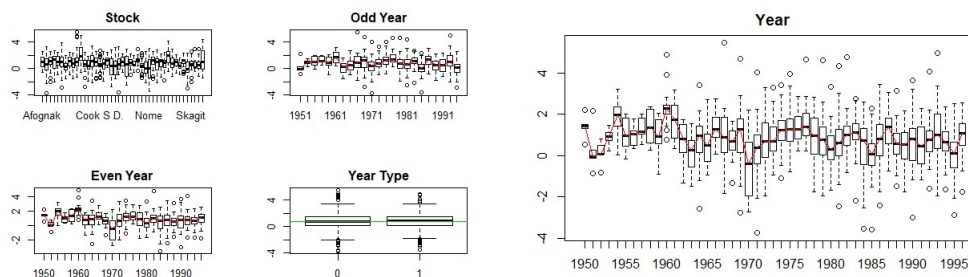
1.5 Graphical Exploration

Below pictures are density distribution for some ordinal/continuous variables in the dataset “pinkdata”.



One can tell that variable year is following a skewed bell shape distribution. Along shore distance, longitude and “gcd” have two peaks. Latitude has multiple peaks.

Below are some boxplots of dependent variable Y, using some categorical variables to separate:

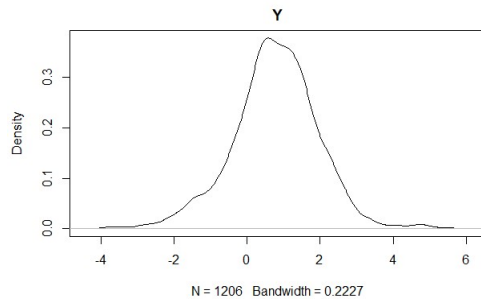


From the boxplot of stock, one can see there are some variations of Y among different stocks.

From the boxplot of odd year and even year, one can see that median of Y does not change very significantly over years. But, by observing these two plots' y-axis, one may see they have different ranges. One can see that, for odd year, the y-axis ranges from -4 to 4 and each year's median value is around 0. For the even year, the y-axis ranges from -2 to 4 and each year's median value is also around 0. From the boxplot of every year together (titled “Year”), one may see some variation for median of Y among years. (Red lines cross medians of each year)

From the boxplot of “year type”, one may see that Y’s median for even year (label 0) is slightly higher than Y’s median for even year (label 1). It is not very significant though. (The green line is the median of Y for odd years, one can see that it is around the bottom of even year’s median broad line)

Below is a density plot of variable of interest, Y:



One can see it is roughly following a symmetric bell shape distribution. One could say it’s roughly following a normal distribution.

1.6 Potential of Addressing Problem

The report will separate all the variables into two parts: geographical variable and timing variable.

Geographical variable includes stock, latitude, longitude, gcd, along shore distance. Indeed, those variables probably have a lot of multicollinearity issues. So, in the model selection, only one or very a few of them should be included in the model. For instance, gcd is computed through latitude and longitude; it also shares a lot similarity with along shore distance. So, if the model includes gcd, it is reasonable to consider the other variables are not necessary.

Timing variable includes year and year type. According to the final project instruction file, year type could be a reasonable variable to be included, since it differentiates pink salmon species. The boxplots of odd year and even year might also suggest “Year Type” could be a useful factor to describe Y’s pattern. One should notice that, using boxplots alone, this is a rather weak statement. But the report will include it first anyway because it could to be an important preliminary knowledge from project instruction.

The report considers it is not reasonable to include Year itself as a factor to describe the model. With limited knowledge about fisheries, I think it is unreasonable to think ratio of recruitment and spawns would follow certain pattern over year. The boxplots also suggest the Y value is only changing randomly over year.

2 Model Selection and Results

This report will use Gibbs sampling to make inference on potential candidate models, and it will be implemented by RJAGS. Details of algorithm will be discussed in each subsection.

2.1 Model 1

The first model would be a simple Hierarchical Linear Regression Model with mixed effects. The report thinks the even/odd year should have a fixed effect on Y because of the sepcies difference, and each stock location has a random effect on Y because of the geographical influences.

2.1.1 Model Set Up

Let $Y = \{Y_{ij}: i = 1, 2, \dots, n_j, \text{ and } j = 1, 2, \dots, 43\}$, where j is the number of stock, n_j is the number of observation in each $stock_j$. Y is the variable of interest, $\log(\text{recruitment/spawns})$.

The report assumes that conditioning on $\alpha, \beta_0^{fish}, \beta_1^{fish}, \tau^2$, then Y_{ij} are mutually independent.

Hence, the likelihood is:

$$f(Y|\alpha, \beta^{fish}, \tau^2) = \prod_{j=1}^J \prod_{i=1}^{n_j} N(y_{ij}; \mu_{ij}^{fish}, \tau^2)$$

where

$$\mu_{ij}^{fish} = \beta_0^{fish} + \beta_1^{fish} * YearType_{ij} + \alpha_j$$

$$\text{assume } f(\alpha|\tau_\alpha^2) = \prod_{j=1}^J N(\alpha_j; 0, \tau_\alpha^2)$$

$$\beta_0^{fish} \sim N(0, 10^{10})$$

$$\beta_1^{fish} \sim N(0, 10^{10})$$

$$P(\tau^2) = IG(0.001, 0.001)$$

$$P(\tau_\alpha^2) = IG(0.001, 0.001)$$

2.1.2 Parameter Interpretation

α_j : the deviation from the mean of Y by different stock location

β_0^{fish} : the overall mean of Y

β_1^{fish} : the effect of even year returning salmon on mean of Y

τ^2 : the overall variation in Y about the mean

τ_α^2 : the variation in stock variation

2.1.3 Assumptions

The first assumption is $Y|(\text{all the parameters})$ are independent. This could be wrong because the data are, in fact, from a time series. For each stock, one year's Y is very likely to associated with previous year's Y . But, such problem is not the core of this course so I would move on and force this assumption to be true.

Another assumption is that $Y_{ij} \sim N(\mu_{ij}^{fish}, \tau^2)$ and $\alpha_j \sim N(0, \tau_\alpha^2)$. Due to the limit length, the report cannot check those assumption rigorously. But, given the fact that the data set has large number of observations, and the variable of interest Y is log scaled, the report will use central limit theory and consider the normality holds anyway.

2.1.4 Prior Distribution Reasoning

For β_0^{fish} , according to the interpretation, there's no reason to give them a very informative prior. Before seeing the data, one cannot say the mean of Y has any special pattern except it's a real number. If it's negative, it means the mean of recruitment/spawn is less than 1; if it's larger than 0, it means the mean of recruitment/spawn is smaller than 1.

Similarly for the β_1^{fish} , there's no reason to think the effect of even year returning salmon must follow some pattern before seeing the data.

So, it is reasonable to make their prior to be $N(0, 10^{10})$, which is very vague.

For τ^2 and τ_α^2 , it is clear that those two values should be larger than 0. But, without any preliminary knowledge and seeing the data, there's no reason to have any more information about them. So, make them to be *Inverse – Gamma*(0.001, 0.001) is a reasonable choice.

2.1.5 Sampling Algorithm

To perform Gibbs Sampling, parameters $\beta_0^{fish}, \beta_1^{fish}, \alpha, \tau^2, \tau_\alpha^2$ must be assigned to initial values. In my codes, both β^{fish} are assigned to be 0, all α_j are assigned to be 0. τ^2 and τ_α^2 are assigned to be 1. There's no particular reason these values are assigned.

It is unclear how RJAGS will sample which parameter first, but below statement is an example of general implementation of how Gibbs sampling works in this specific case:

We want to sample from $[\beta_0^{fish}, \beta_1^{fish}, \alpha, \tau^2, \tau_\alpha^2 | X]$, we have $\beta_0^{fish}[0], \beta_1^{fish}[0], \alpha[0], \tau^2[0], \tau_\alpha^2[0]$. Then, for $j = 1, 2, \dots$ repeat these k steps:

generate $\beta_0^{fish}[j] \sim [\beta_0^{fish} | \beta_1^{fish}[j-1], \alpha[j-1], \tau^2[j-1], \tau_\alpha^2[j-1]]$

generate $\beta_1^{fish}[j] \sim [\beta_1^{fish} | \beta_0^{fish}[j-1], \alpha[j-1], \tau^2[j-1], \tau_\alpha^2[j-1]]$

generate $\alpha[j] \sim [\alpha | \beta_0^{fish}[j-1], \beta_1^{fish}[j-1], \tau^2[j-1], \tau_\alpha^2[j-1]]$

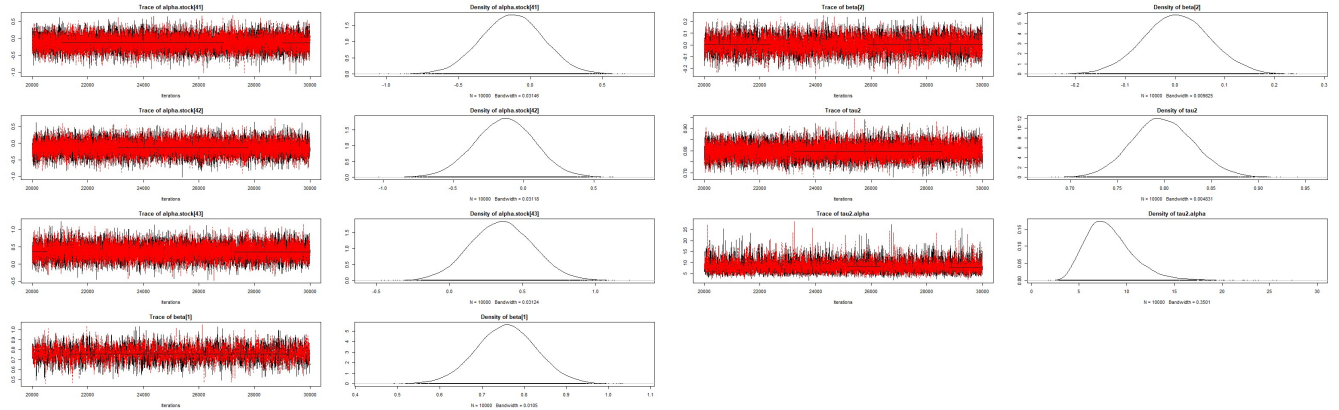
generate $\tau^2[j] \sim [\tau^2 | \beta_0^{fish}[j-1], \beta_1^{fish}[j-1], \alpha[j-1], \tau_\alpha^2[j-1]]$

generate $\tau_\alpha^2[j] \sim [\tau_\alpha^2 | \beta_0^{fish}[j-1], \beta_1^{fish}[j-1], \alpha[j-1], \tau^2[j-1]]$

The order of sampling which parameter can be different. It is unclear how RJAGS's sampler work but above generating distributions are full conditional distribution, in theory.

2.1.6 Results

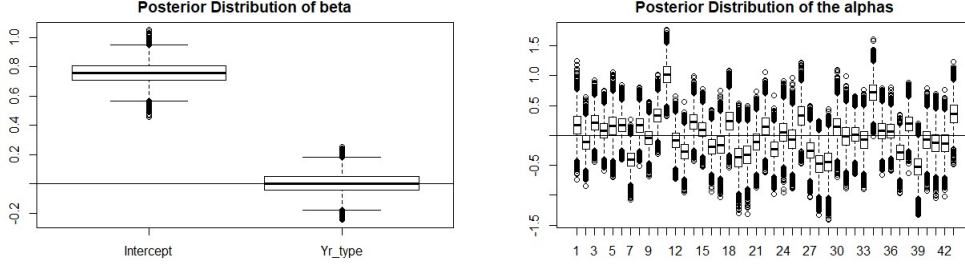
All parameters converged well in this model, below are two trace plots for demonstration. Including all trace plots will take too much space. Anyone who's interested in seeing all trace plots can refer to the Appendix2: Model1's Trace Plot.



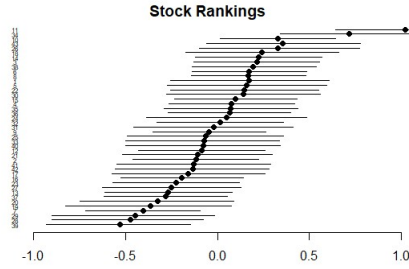
Below are boxplots for $\beta_0^{fish}, \beta_1^{fish}$ and all α . Those boxplots may used to describe those parameters posterior distribution.

One can see that Intercept, β_0 , has all samples larger than 0, which could imply that overall mean of Y is larger than 0. So, overall, we may conclude that recruit's number is larger than spawn's. Because it is hard to set seed for RJAGS, the report decides to not include any confidence interval or significant test in this case since the result is so significant by just observing. One may also see that β_1^{fish} 's median is at 0 and its other simulated values are evenly distributed around 0. It means salmons return at even year and odd year don't have much difference on recruit/spawns ratio. Again, confidence interval will not be provided

since the plot is very significant clear and the numerical result will be different every time because of no setting seed.



One can also see that for different stock locations, the distributions of α are very different, below's posterior expected ranking plot can give a better demonstration.



From above stock ranking plot, one can easily see that α values at different locations have a linear trend of increase. This may provide us some more insight on improving the model.

2.2 Model 2

From model 1, the report finds that even/odd year may not be a significant factor. Also, it finds that different stock locations have very different parameter values. So, it is reasonable to consider drop the variable β_1^{fish} and incorporating stock location covariates.

The report decided to use “gcd” as new variable for stock location’s characteristic. Latitude, longitude and along shore distance are all variables can be considered characters for stock location. But, “gcd” is directly computed through latitude and longitude. Also, from EDA and instruction file, along shore distance and “gcd” are very similar variables. So, I believe including one of them is sufficient enough.

Further, using resources file, one can easily know what is “gcd” exactly. Yet, how along shore distance was measured is unknown.

2.2.1 Model Set Up

Let $Y = \{Y_{ij}: i = 1, 2, \dots, n_j, \text{ and } j = 1, 2, \dots, 43\}$, where j is the number of stock, n_j is the number of observation in each $stock_j$. Y is the variable of interest, $\log(\text{recruitment/spawns})$.

The report assumes that conditioning on $\alpha, \beta_0^{fish}, \beta_1^{fish}, \tau^2$, then Y_{ij} are mutually independent.

Hence, the likelihood is:

$$f(Y|\alpha, \beta^{fish}, \tau^2) = \prod_{j=1}^J \prod_{i=1}^{n_j} N(y_{ij}; \mu_{ij}^{fish}, \tau^2)$$

where

$$\mu_{ij}^{fish} = \beta_0^{fish} + \alpha_j$$

$$\text{assume } f(\alpha|\tau_\alpha^2) = \prod_{j=1}^J N(\alpha_j; \mu_j^{stock}, \tau_\alpha^2)$$

$$\mu_j^{stock} = \beta^{stock} * gcd_j$$

$$\beta_0^{fish} \sim N(0, 10^{10})$$

$$\beta^{stock} \sim N(0, 10^{10})$$

$$P(\tau^2) = IG(0.001, 0.001)$$

$$P(\tau_\alpha^2) = IG(0.001, 0.001)$$

The second model is largely same with first model, except that α 's mean will not be 0 but μ^{stock} , which is in a linear relationship with “gcd”.

2.2.2 Parameter Interpretation

α_j : the deviation from the mean of Y by different stock location

β_0^{fish} : the overall mean of Y

β^{stock} : the effect of one unit increase of “gcd” on mean of Y

τ^2 : the overall variation in Y about the mean

τ_α^2 : the variation in stock variation

2.2.3 Assumption

The model assumptions are very similar to model 1. Except that $\alpha_j \sim N(\mu^{stock}, \tau_\alpha^2)$ instead of $\alpha_j \sim N(0, \tau_\alpha^2)$.

Also, it requires that $\pi(\beta^{fish}, \beta^{stock}, \tau^2, \tau_\alpha^2) = \pi(\beta^{fish})\pi(\beta^{stock})\pi(\tau^2)\pi(\tau_\alpha^2)$

2.2.4 Prior Distribution Reasoning

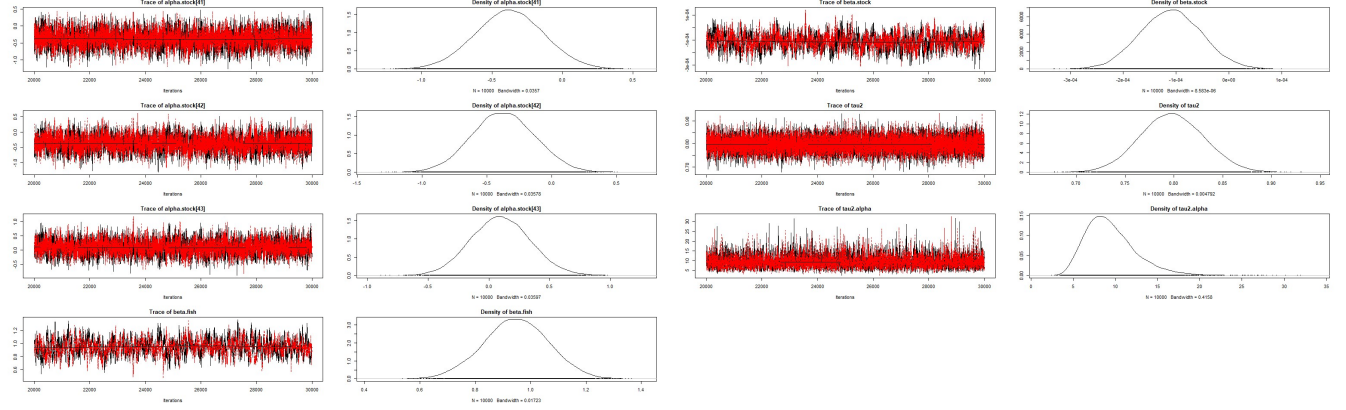
The new parameter is β^{stock} and its prior distribution is assumed to be $N(0, 10^{10})$ again. The reason is similar with before, without seeing the data and preliminary knowledge, there's no reason to believe the effect of one unit increase of “gcd” on mean of Y should has any special pattern except it's a real number.

2.2.5 Sampling Algorithm

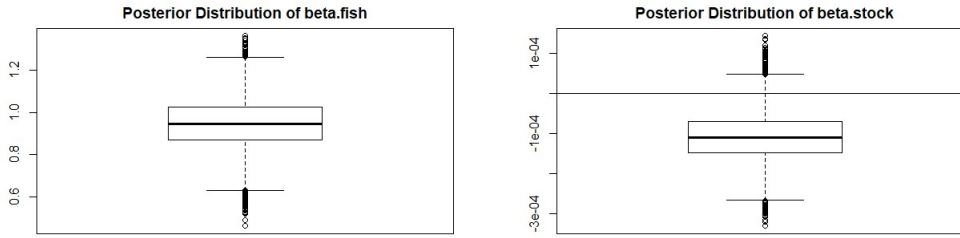
The algorithm will be very similar to the model1. The report will not state it again.

2.2.6 Results

All parameters converged well in this model. Again, the report will only include two trace plots for demo. Please refer to Appendix3: Model2 Trace Plots for full trace plot.



Below are boxplots for β_0^{fish} and β^{stock} .



Once again, the β_0^{fish} has no values lower than 0. The result is very significant and the report will not carry out a numerical confidence interval for it.

However, the boxplot for β^{stock} may not be very clear. So, the report will carry out a confidence interval for the parameter. But do notice that, there's no seed set for this so result may vary from time to time. However, it should give a general idea about significant level of the parameter.

Using 90% confidence interval, the confidence interval of β^{stock} is:

$$(-0.0001097452 - 1.645 * 5.868554 * 10^{-5}, -0.0001097452 + 1.645 * 5.868554 * 10^{-5})$$

$$\text{or } (-0.0002062829, -1.32075 * 10^{-5})$$

which does not contain 0, one may conclude that with 90% confidence level, one unit increase of “gcd” will have negative effect on Y.

This makes sense; if the “gcd” is larger, the stock location is more far away from the sea. Those spawns from such location will confront themselves with more predators when they swimming back. More of them will get killed and less recruits will be hatched.

One may also noticed that the scale of the parameter β^{stock} is very small. The reason is because the “gcd” has a much larger data scale than Y. This should not be considered as potential problem for not significant. Small coefficient does not imply non-significant.

3. Model Selection

The model selection will take advantage of DIC.

Using the R command `dic.samples` and `diffdic`, it indicates that model1's DIC is larger, so model2 is a better fit.

This is consistent with the report's previous idea, Model2 fits the data better. However, the computation and details of DIC is unclear to me at this point of time. So, this may not be a 100% accurate statement.

4. Discussion and Conclusion

In conclusion, the $\log(\text{Recruitment}/\text{Spawn})$ has a negative linear association with great circle distance from the most south stock. In another word, salmons from stock far away from the ocean tend to have a lower $\log(\text{Recruitment}/\text{Spawn})$. This is not surprising since salmons from such stocks face more dangerous incidents when they go back to their birth location and reproduce.

However, even though salmons return by even years and odd years are different species, their ratio of $\text{Recruitment}/\text{Spawn}$ is the same. It appears that return by even year or odd year does not have an effect on ratio of $\text{Recruitment}/\text{Spawn}$.

This report assumed that each observation is independent to another, which is most likely not true. The data are from a time series data, so it is reasonable to at least consider observation for one stock from different years are correlated.

This report also assumed normality of Y and effect of stock at each stock location. This may not be true and may cause problems.

Another assumption deserve to be noticed is that for variance parameters, the prior distributions of Inverse Gamma may not be so non-informative. This report chooses it is because it will cover a very large range of positive real number. A further step could be using sensitive analysis to take a look other reasoanbel prior.

Appendix

1 Full Codes

Loading of Data

```
# read in .csv data
data<-read.csv('pinkdata.csv')
# add vector of observation numbers
data[['obs']]<-seq(1,dim(data)[1])
# calculate log recruits to spawners variable and add to list
data[['Y']]<-log(data$R/data$S)
# create a new variable X and give it the values in S
data[['X']]<-data$S
# create a new variable yr and give it the values in BY
data[['yr']]<-data$BY
# create a new variable asd and give it the values in AlongShore_Distance
data[['asd']]<-data$AlongShore_Distance
# convert list to data frame
df.pink<-data.frame(data)
# remove missing observations
missing<-which(is.nan(df.pink[, 'Y']))
if (length(missing)>0) {df.pink<-df.pink[-missing,]}
```

function to calculate Great-Circle distance between two stocks

```
gcdcalc<-function(lon1,lat1,lon2,lat2)
{
  r<-6378
  gcd<-r*acos(sin(lat1*pi/180)*sin(lat2*pi/180)+cos(lat1*pi/180)*cos(lat2*pi/180)*cos((lon1-lon2)*pi/180))
  gcd
}

# find southernmost stock location
zerogcd<-min(df.pink[, 'Latitude'], na.rm=TRUE)
smost.ind<-which(df.pink[, 'Latitude']==zerogcd)[1]
# measure Great Circle Distance to all other stock locations from Southernmost stock
df.pink[, 'gcd']<-gcdcalc(df.pink[smost.ind, 'Longitude'], df.pink[smost.ind, 'Latitude'], df.pink[, 'Longitude'])

# get column names
#attributes(df.pink)$names
```

Missing Data

```
## Take a look of other missing values
## below are all the independent original data
sum(is.na(df.pink$Species))
sum(is.na(df.pink$Stock))
```

```

sum(is.na(df.pink$Region))
sum(is.na(df.pink$Latitude))
sum(is.na(df.pink$Longitude))
sum(is.na(df.pink$Longitude))
sum(is.na(df.pink$AlongShore_Distance))
sum(is.na(df.pink$StockNumber))
sum(is.na(df.pink$BY))

missing_obs <- df.pink[is.na(df.pink$Latitude),]
#missing_obs

## Not missing by random

## check if there's "NA" characters
#df.pink <- na.omit(df.pink)
##sum(df.pink == 'NA')

## adding even and odd year as a variable
df.pink[, "Yr_Type"] <- as.numeric(df.pink$yr%%2 == 0)
#df.pink$yr <- as.factor(df.pink$yr)
attach(df.pink)

```

EDA of df.pink

```

par(mfrow=c(2,2))

d.year <- density(df.pink$yr)
plot(d.year, main = "Year")

d.asd <- density(df.pink$asd, na.rm = T)
plot(d.asd, main = "ASD")

d.lat <- density(df.pink$Latitude, na.rm = T)
plot(d.lat, main = "Latitude")

d.lon <- density(df.pink$Longitude, na.rm = T)
plot(d.lon, main = "Longitude")

par(mfrow=c(1,1))
d.gcd <- density(df.pink$gcd, na.rm = T)
plot(d.gcd, main = "gcd")

```

continue EDA

```

par(mfrow=c(2,2))
#boxplot(df.pink$Y ~ df.pink$Region)
boxplot(df.pink$Y ~ df.pink$Stock, main = "Stock")

year_m <- aggregate(Y ~ yr, data = df.pink, median)

```

```

boxplot(df.pink[Yr_Type == 0, ]$Y ~ df.pink[Yr_Type == 0,]$yr, main = "Odd Year")
lines(1:23, year_m$Y[seq(2, 47, 2)], col = 2)

boxplot(df.pink[Yr_Type == 1, ]$Y ~ df.pink[Yr_Type == 1,]$yr, main = "Even Year")
lines(1:24, year_m$Y[seq(1, 47, 2)], col = 2)

boxplot(df.pink$Y ~ Yr_Type, main = "Year Type")
abline(h = median(df.pink[Yr_Type == 0, ]$Y), col = 3)

par(mfrow=c(1,1))
boxplot(df.pink$Y ~ yr, main = "Year")

lines(1:47, year_m$Y, col = 2)

```

```

d.Y <- density(df.pink$Y)
plot(d.Y, main = "Y")

```

```

library(rjags)
N.stock <- length(unique(Stock))

N.Y <- length(unique(Y))

P.fish <- 2

Datalist <- list(
  "Y" = Y,
  "Yr_Type" = Yr_Type,
  "Stock" = Stock,
  "N.Y" = N.Y,
  "P.fish" = P.fish,
  "N.stock" = N.stock
)

# List of parameters to be monitored
parameters <- c(
  "beta",
  "tau2",
  "alpha.stock",
  "tau2.alpha"
)

# Set initial values
initsValues <- list(
  "beta" = rep(0, P.fish),
  "tau2" = 1,
  "alpha.stock" = rep(0, N.stock),
  "tau2.alpha" = 1
)

# JAGS Set-up
adaptSteps <- 20000          #number of steps to "tune" the samplers
burnInSteps <- 20000        #number of steps to "burn-in" the samplers
nChains <- 2                 #number of chains to run

```

```

numSavedSteps <- 20000          #total number of steps in chains to save
thinSteps <- 1                 #number of steps to "thin" (1=keep every step)
nIter <- ceiling((numSavedSteps*thinSteps)/nChains) #steps per chain

# Create, initialize, and adapt the model
jagsModel <- jags.model("fishery_model1.txt",
                      data=Datalist,
                      inits=initsValues,
                      n.chains=nChains,
                      n.adapt=adaptSteps)

update(jagsModel, n.iter = burnInSteps)

codaSamples <- coda.samples(jagsModel,
                           variable.names = parameters,
                           n.iter = nIter,
                           thin = thinSteps)

par(ask=F)
plot(codaSamples)

#traceplot(codaSamples)

# Retrieve posterior samples
mcmcChain <- as.matrix(codaSamples)
betaSamples <- matrix(NA, numSavedSteps, P.fish)

for(i in 1:P.fish) betaSamples[,i] <- mcmcChain[, paste("beta[",i,"]", sep="")]

par(mfrow=c(1,1), ask=F)
boxplot(as.data.frame(betaSamples),
        names=c("Intercept","Yr_type"),
        main="Posterior Distribution of beta")
abline(h=0)

# Examine the posterior distribution of the stock random effect
alphaSamples <- matrix(NA, numSavedSteps, N.stock)
for(j in 1:N.stock) alphaSamples[,j] <- mcmcChain[, paste("alpha.stock[",j,"]", sep="")]

par(mfrow=c(1,1), ask=F)
boxplot(as.data.frame(alphaSamples),
        names=as.character(1:N.stock),
        main="Posterior Distribution of the alphas")
abline(h=0)

rank.alpha <- matrix(NA, numSavedSteps, N.stock)
for(k in 1:numSavedSteps){
  rank.alpha[k,] <- rank(alphaSamples[k,])
}
avg.rank <- rank(apply(rank.alpha, 2, mean))

par(mfrow=c(1,1), ask=F)
plot(c(-1.1,1), c(1,42), type='n', axes=F, xlab="", ylab="", main="Stock Rankings")

```

```

axis(1,at=seq(-1, 1, length=5))
for(j in 1:N.stock)
{
  lines(quantile(alphaSamples[,j], c(0.025,0.975)), rep(avg.rank[j],2))
  points(mean(alphaSamples[,j]), avg.rank[j], pch=19)
  text(-1.1, avg.rank[j], j, cex=.5)
}

```

Second Model

```

df.stock <- read.csv("stockinfo.csv")
df.stock <- df.stock[df.stock$Species == 'pink',]
#df.stock <- df.stock[df.stock$Stock != 'Puyallup',] #missing data eliminated

# find southernmost stock location
zerogcd<-min(df.stock[, 'Latitude'], na.rm=TRUE)
smost.ind<-which(df.stock[, 'Latitude']==zerogcd)[1]
# measure Great Circle Distance to all other stock locations from Southernmost stock
df.stock[, 'gcd']<-gcdcalc(df.stock[smost.ind, 'Longitude'], df.stock[smost.ind, 'Latitude'], df.stock[, 'Longitude'], df.stock[, 'Latitude'])

attach(df.stock)

```

Second Model

```

P.stock <- 1
N.stock <- length(unique(df.stock$Stock))
P.fish <- 1

Datalist <- list(
  "Y" = Y,
  #"Yr_Type" = Yr_Type,
  "Stock" = df.pink$Stock,
  "N.Y" = N.Y,
  "P.fish" = P.fish,
  "P.stock" = P.stock,
  "N.stock" = N.stock,
  "gcd" = gcd
  #"asd" = AlongShore.Distance
)

# List of parameters to be monitored
parameters <- c(
  "beta.fish",
  "beta.stock",
  "tau2",
  "alpha.stock",
  "tau2.alpha"
)

```

```

# Set initial values
initsValues <- list(
  "beta.fish" = rep(0, P.fish),
  "beta.stock" = rep(0, P.stock),
  "tau2" = 1,
  "alpha.stock" = rep(0, N.stock),
  "tau2.alpha" = 1
)

jagsModel3 <- jags.model("fishery_model3.txt",
  data=Datalist,
  inits=initsValues,
  n.chains=nChains,
  n.adapt=adaptSteps)

update(jagsModel3, n.iter = burnInSteps)

codaSamples3 <- coda.samples(jagsModel3,
  variable.names = parameters,
  n.iter = nIter,
  thin = thinSteps)

par(ask=F)
plot(codaSamples3)

# Retrieve posterior samples
mcmcChain3 <- as.matrix(codaSamples3)
betaSamples3 <- matrix(NA, numSavedSteps, P.fish)

for(i in 1:P.fish) betaSamples3[,i] <- mcmcChain3[, paste("beta.fish", sep="")]

par(mfrow=c(1,1), ask=F)
boxplot(as.data.frame(betaSamples3),
  names=c("Intercept"),
  main="Posterior Distribution of beta.fish")
abline(h=0)

betaSamples3 <- matrix(NA, numSavedSteps, P.stock)
#for(i in 1:P.stock) betaSamples2[,i] <- mcmcChain2[, paste("beta.stock[",i,"]", sep="")]

for(i in 1:P.stock) betaSamples3[,i] <- mcmcChain3[, paste("beta.stock", sep="")]

par(mfrow=c(1,1), ask=F)
boxplot(as.data.frame(betaSamples3), names=c("gcd"), main="Posterior Distribution of beta.stock")
abline(h=0)

# confidence interval
mcmcChain3 <- as.data.frame(mcmcChain3)
mean(mcmcChain3$beta.stock)
sd(mcmcChain3$beta.stock)
mean(mcmcChain3$beta.stock) - 1.645*sd(mcmcChain3$beta.stock)
mean(mcmcChain3$beta.stock) + 1.645*sd(mcmcChain3$beta.stock)

dic1 <- dic.samples(jagsModel, nIter)
dic3 <- dic.samples(jagsModel3, nIter)

```

```
diffdic(dic1, dic2)
diffdic(dic2, dic3)
diffdic(dic1, dic3)
```