Black/Asian carp model selection

Eddie Wu

2023-08-02

Introduction

This .Rmd file is to show the progress on black carp and asian carp temperature and condition analyses. Since sub-sampling from spatial autocorrelation does not give significantly different results from normal analysis for black carp, we present the results without sub-sampling here. For other asian carp species, we still subsample.

For each species, we have four different models:

1. Simple linear model (same slope, same intercept)
2. Linear additive model (same slope, different intercept)
3. Interaction model (different slope, same intercept)
4. Group-specific model (different slope, different intercept)

And we consider two temperature metrics:

1. Annual temperature
2. Winter temperature (temperature from the coldest quarter)

```
library(ggplot2)
library(ggfortify)
library(dplyr)
library(knitr)

## Import data
asian.carp <- read.csv("asian carp final.csv")
asian.carp$Condition <- as.factor(asian.carp$Condition)

Black <- read.csv("eddie_carp_new.csv")
Black$condition <- as.factor(Black$condition)

## Separate by species
Grass <- asian.carp[asian.carp$Species=="Grass",]
Bighead <- asian.carp[asian.carp$Species=="Bighead",]
Silver <- asian.carp[asian.carp$Species=="Silver",]
Big.sil <- rbind(Bighead, Silver) # combine the two groups
```

Black carp

For black carp data, we do not subsample since sub-sampling at 250 km does not reduce our moran's I.

Using annual temperature

```
# Clean data
black.clean <- Black %>% filter(!row_number() == 5) %>% filter(sex != "male")
```

```r
# Build the models
black.simple <- lm(log(AAM)~AnnualTemp, data = black.clean)
black.linear <- lm(log(AAM)~AnnualTemp+condition, data = black.clean)
black.int <- lm(log(AAM)~AnnualTemp:condition, data = black.clean)
black.group <- lm(log(AAM)~AnnualTemp*condition, data = black.clean)

# Compare the AICs
AIC(black.simple, black.linear, black.int, black.group)
```

```
##              df        AIC
## black.simple  3 -2.5751205
## black.linear  4 -1.8897277
## black.int     4 -2.7821124
## black.group   5 -0.8525415
```

```r
# R^2 value for the four models
r_2 <- data.frame(
  Model = c("Simple linear", "Linear additive", "Interaction", "Grouped"),
  R2 = c(summary(black.simple)$r.squared, summary(black.linear)$r.squared,
         summary(black.int)$r.squared, summary(black.group)$r.squared)
)
kable(r_2)
```

Model	R2
Simple linear	0.2873035
Linear additive	0.3268967
Interaction	0.3525125
Grouped	0.3544922

Using cold temperature

```r
# Build the models
black.simple <- lm(log(AAM)~ColdTemp, data = black.clean)
black.linear <- lm(log(AAM)~ColdTemp+condition, data = black.clean)
black.int <- lm(log(AAM)~ColdTemp:condition, data = black.clean)
black.group <- lm(log(AAM)~ColdTemp*condition, data = black.clean)

# Compare the AICs
AIC(black.simple, black.linear, black.int, black.group)
```

```
##              df       AIC
## black.simple  3 -3.257277
## black.linear  4 -2.574254
## black.int     4 -2.154127
## black.group   5 -1.758091
```

```r
# R^2 value for the four models
r_2 <- data.frame(
  Model = c("Simple linear", "Linear additive", "Interaction", "Grouped"),
  R2 = c(summary(black.simple)$r.squared, summary(black.linear)$r.squared,
         summary(black.int)$r.squared, summary(black.group)$r.squared)
)
kable(r_2)
```

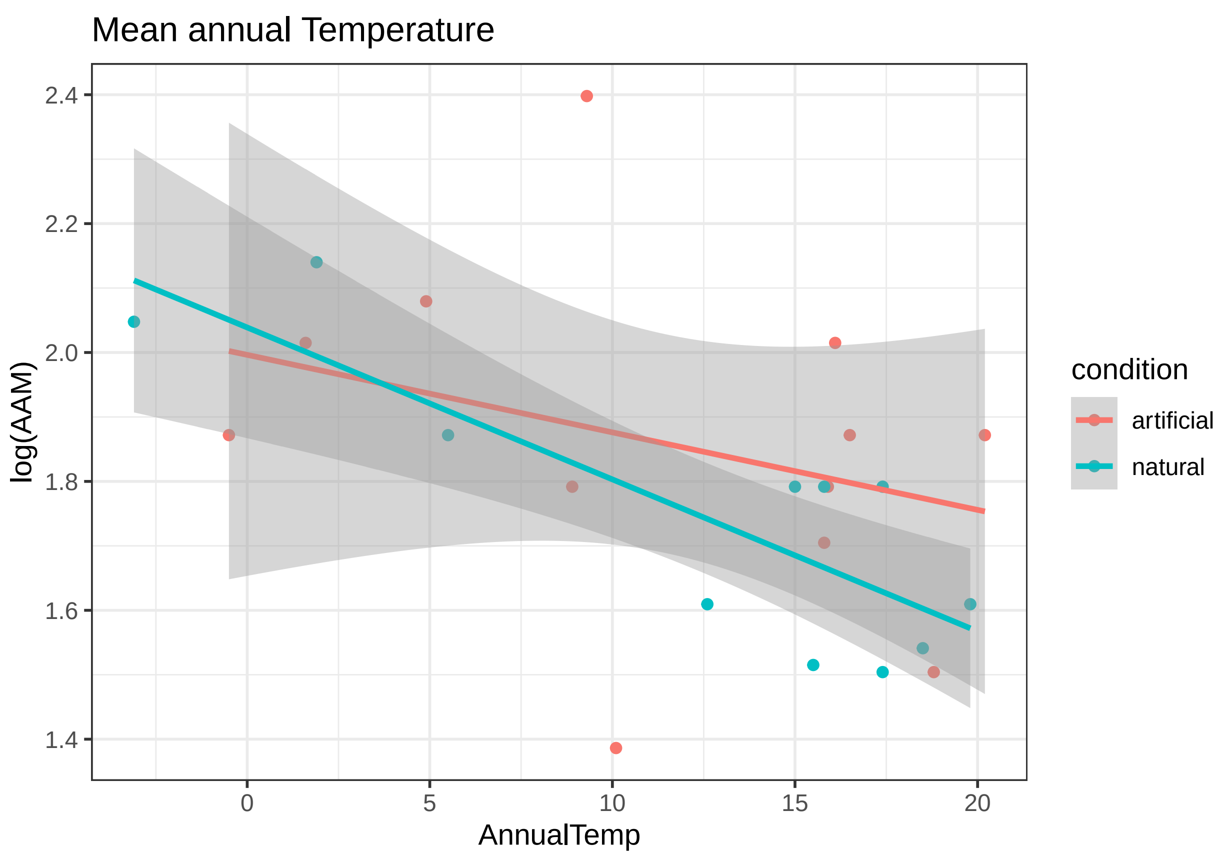
Model	R2
Simple linear	0.3081310
Linear additive	0.3466344
Interaction	0.3345901
Grouped	0.3794131

We separated the black carp dataset into two based on conditions. Since there was no preference over the four models, we used the simple linear model on each set of the data.

Black carp graphs with two conditions separated

```
## Annual temperature
ggplot(black.clean, aes(x = AnnualTemp, y = log(AAM), color = condition))+
  geom_point()+
  geom_smooth(method = "lm")+
  theme_bw()+
  labs(title = "Mean annual Temperature")
```

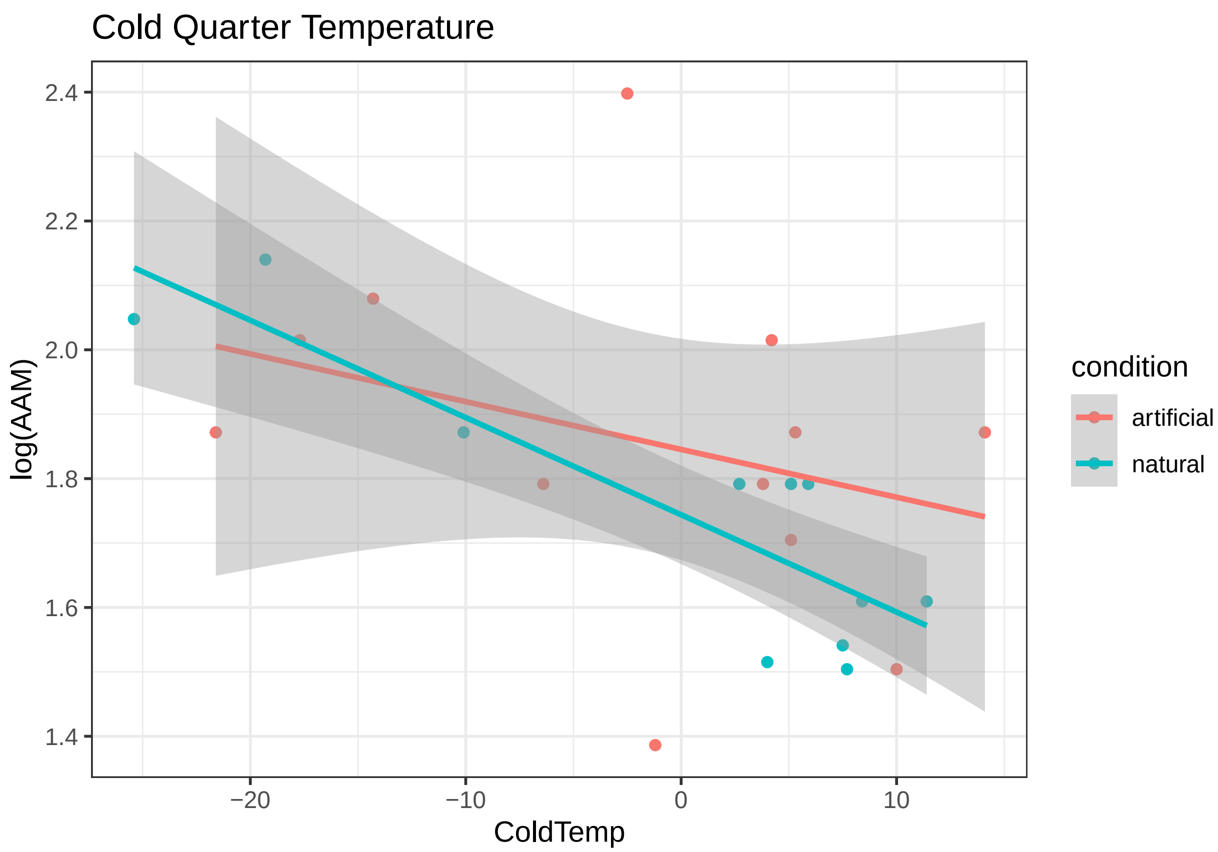
```
## `geom_smooth()` using formula 'y ~ x'
```




```
## Cold temperature
ggplot(black.clean, aes(x = ColdTemp, y = log(AAM), color = condition))+
  geom_point()+
  geom_smooth(method = "lm")+
  theme_bw()+
```

```r
  labs(title = "Cold Quarter Temperature")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



Now that we have seen that the artificial condition data seems to have a larger spread, we would like to run the simple linear model to take a look.

```r
## Separate into two data sets
black.natural <- black.clean[black.clean$condition == "natural",]
black.artificial <- black.clean[black.clean$condition == "artificial",]


## Run the models
black.annual.n <- lm(log(AAM)~AnnualTemp, data = black.natural)
black.cold.n <- lm(log(AAM)~ColdTemp, data = black.natural)

black.annual.a <- lm(log(AAM)~AnnualTemp, data = black.artificial)
black.cold.a <- lm(log(AAM)~ColdTemp, data = black.artificial)


## Compare the AIC scores
AIC(black.annual.n, black.annual.a) #for annual temperature
```

```
## Warning in AIC.default(black.annual.n, black.annual.a): models are not all
## fitted to the same number of observations
```

```
##                 df       AIC
## black.annual.n  3 -10.49220
```

```
## black.annual.a  3   5.89183
```

```r
AIC(black.cold.n, black.cold.a) #for cold temperature
```

```
## Warning in AIC.default(black.cold.n, black.cold.a): models are not all fitted
## to the same number of observations

##              df       AIC
## black.cold.n  3 -13.12105
## black.cold.a  3   5.85967
```

```r
## Compare the model parameters
summary(black.annual.n)
```

```
##
## Call:
## lm(formula = log(AAM) ~ AnnualTemp, data = black.natural)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.15831 -0.09440 -0.03738  0.11596  0.16311
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.038839   0.075970  26.838  6.7e-10 ***
## AnnualTemp  -0.023574   0.005304  -4.445  0.00161 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1264 on 9 degrees of freedom
## Multiple R-squared:  0.687,  Adjusted R-squared:  0.6523
## F-statistic: 19.76 on 1 and 9 DF,  p-value: 0.001612
```

```r
summary(black.annual.a)
```

```
##
## Call:
## lm(formula = log(AAM) ~ AnnualTemp, data = black.artificial)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.48857 -0.10881  0.01223  0.12431  0.51341
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.99633    0.15388  12.973  1.4e-07 ***
## AnnualTemp  -0.01203    0.01166  -1.031    0.327
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2639 on 10 degrees of freedom
## Multiple R-squared:  0.09614,    Adjusted R-squared:  0.005759
## F-statistic: 1.064 on 1 and 10 DF,  p-value: 0.3267
```

```r
summary(black.cold.n)
```

```
##
```

```
## Call:
## lm(formula = log(AAM) ~ ColdTemp, data = black.natural)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.168342 -0.084535 -0.007609  0.096764  0.136973
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.743853   0.033825  51.555 1.95e-12 ***
## ColdTemp    -0.015096   0.002878  -5.246 0.000531 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1122 on 9 degrees of freedom
## Multiple R-squared:  0.7536, Adjusted R-squared:  0.7262
## F-statistic: 27.52 on 1 and 9 DF,  p-value: 0.0005305
```

```r
summary(black.cold.a)
```

```
##
## Call:
## lm(formula = log(AAM) ~ ColdTemp, data = black.artificial)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.46794 -0.11050  0.00646  0.12880  0.53402
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.845330   0.077097  23.935 3.68e-10 ***
## ColdTemp    -0.007416   0.007092  -1.046     0.32
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2635 on 10 degrees of freedom
## Multiple R-squared:  0.09856,    Adjusted R-squared:  0.00842
## F-statistic: 1.093 on 1 and 10 DF,  p-value: 0.3203
```

It turned put that after separating out the artificial condition, the model performed much better. While the artificial model alone did not even have a significant relationship.

Asian carp

Using annual temperature

```r
## Look at the spatial codes for the current asian carp data
asian.carp.clean <- asian.carp %>%
  filter(Condition %in% c("natural", "artificial"))

table(asian.carp.clean$Code)
```

```
##
##  A AA AB AC AD AE AF AG AH AI AL AM AN AO AP  B  C  D  E  F  G  H  J  K  L  M
##  7  1  1  1  1  1  3  3  1  1  2  1  1  1  2  6  3  4  3  1  2  1  2  1  1  3
##  N  O  Q  S  X  Y  Z
```

```
## 1 3 1 3 1 1 3
```

```r
## Simple linear model
asian.linear.results <- matrix(NA,1000,2)
for(i in 1:1000){
  sub <- asian.carp.clean %>% group_by(Code) %>% sample_n(size=1)
  reg <- lm(log(AAM)~AnnualTemp, data = sub)
  asian.linear.results[i,1]<-as.numeric(AIC(reg))
  asian.linear.results[i,2]<-summary(reg)$adj.r.squared
}


## Linear addictive models (same slope, different intercept)
asian.add.results <- matrix(NA,1000,2)
for(i in 1:1000){
  sub <- asian.carp.clean %>% group_by(Code) %>% sample_n(size=1)
  reg <- lm(log(AAM)~AnnualTemp+Condition, data = sub)
  asian.add.results[i,1]<-as.numeric(AIC(reg))
  asian.add.results[i,2]<-summary(reg)$adj.r.squared
}


## Interaction models (different slope, same intercept)
asian.int.results <- matrix(NA,1000,2)
for(i in 1:1000){
  sub <- asian.carp.clean %>% group_by(Code) %>% sample_n(size=1)
  reg <- lm(log(AAM)~AnnualTemp:Condition, data = sub)
  asian.int.results[i,1]<-as.numeric(AIC(reg))
  asian.int.results[i,2]<-summary(reg)$adj.r.squared
}


## Grouped-specific model (different slope, different intercept)
asian.group.results <- matrix(NA,1000,2)
for(i in 1:1000){
  sub <- asian.carp.clean %>% group_by(Code) %>% sample_n(size=1)
  reg <- lm(log(AAM)~AnnualTemp*Condition, data = sub)
  asian.group.results[i,1]<-as.numeric(AIC(reg))
  asian.group.results[i,2]<-summary(reg)$adj.r.squared
}


## Compare the AIC results
mean(unique(asian.linear.results[,1]))
```

```
## [1] 17.35023
```

```r
mean(unique(asian.add.results[,1]))
```

```
## [1] 18.96549
```

```r
mean(unique(asian.int.results[,1]))
```

```
## [1] 18.21879
```

```r
mean(unique(asian.group.results[,1]))
```

```
## [1] 19.94214
```

```r
# R^2 value for the four models
r_2 <- data.frame(
  Model = c("Simple linear", "Linear additive", "Interaction", "Grouped"),
  R2 = c(mean(unique(asian.linear.results[,2])),
         mean(unique(asian.add.results[,2])),
         mean(unique(asian.int.results[,2])),
         mean(unique(asian.group.results[,2])))
)
kable(r_2)
```

Model	R2
Simple linear	0.5486467
Linear additive	0.5385891
Interaction	0.5492742
Grouped	0.5371078

- No preference among all four models.

Using cold temperature

```r
## Simple linear model
asian.linear.results <- matrix(NA,1000,2)
for(i in 1:1000){
  sub <- asian.carp.clean %>% group_by(Code) %>% sample_n(size=1)
  reg <- lm(log(AAM)~ColdTemp, data = sub)
  asian.linear.results[i,1]<-as.numeric(AIC(reg))
  asian.linear.results[i,2]<-summary(reg)$adj.r.squared
}


## Linear addictive models (same slope, different intercept)
asian.add.results <- matrix(NA,1000,2)
for(i in 1:1000){
  sub <- asian.carp.clean %>% group_by(Code) %>% sample_n(size=1)
  reg <- lm(log(AAM)~ColdTemp+Condition, data = sub)
  asian.add.results[i,1]<-as.numeric(AIC(reg))
  asian.add.results[i,2]<-summary(reg)$adj.r.squared
}


## Interaction models (different slope, same intercept)
asian.int.results <- matrix(NA,1000,2)
for(i in 1:1000){
  sub <- asian.carp.clean %>% group_by(Code) %>% sample_n(size=1)
  reg <- lm(log(AAM)~ColdTemp:Condition, data = sub)
  asian.int.results[i,1]<-as.numeric(AIC(reg))
  asian.int.results[i,2]<-summary(reg)$adj.r.squared
}


## Grouped-specific model (different slope, different intercept)
```

```r
asian.group.results <- matrix(NA,1000,2)
for(i in 1:1000){
  sub <- asian.carp.clean %>% group_by(Code) %>% sample_n(size=1)
  reg <- lm(log(AAM)~ColdTemp*Condition, data = sub)
  asian.group.results[i,1]<-as.numeric(AIC(reg))
  asian.group.results[i,2]<-summary(reg)$adj.r.squared
}


## Compare the AIC results
mean(unique(asian.linear.results[,1]))
```

```
## [1] 18.66339
```

```r
mean(unique(asian.add.results[,1]))
```

```
## [1] 19.55866
```

```r
mean(unique(asian.int.results[,1]))
```

```
## [1] 20.34613
```

```r
mean(unique(asian.group.results[,1]))
```

```
## [1] 20.77791
```

```r
# R^2 value for the four models
r_2 <- data.frame(
  Model = c("Simple linear", "Linear additive", "Interaction", "Grouped"),
  R2 = c(mean(unique(asian.linear.results[,2])),
         mean(unique(asian.add.results[,2])),
         mean(unique(asian.int.results[,2])),
         mean(unique(asian.group.results[,2])))
)
kable(r_2)
```

Model	R2
Simple linear	0.5308302
Linear additive	0.5308670
Interaction	0.5193119
Grouped	0.5248670

- Same conclusion when using the cold temperature. But using annual temperature seems to have a better model as explained by lower AIC values.

Grass carp

Using annual temperature

```r
Grass.clean <- Grass %>%
  filter(Condition %in% c("natural", "artificial"))

table(Grass.clean$Code)
```

```
##
##  A AA AB AE AF AG AH AL AM AN AO AP  B  C  D  E  F  G  J  K  L  M  O  Q  S  X
```

```
##   3   1   1   1   1   1   1   1   1   1   1   1   3   2   2   1   1   2   1   1   1   1   1   1   1   1
##   Y   Z
##   1   1
```

```r
## Simple linear model
grass.linear.results <- matrix(NA,100,2)
for(i in 1:100){
  sub <- Grass.clean %>% group_by(Code) %>% sample_n(size=1)
  reg <- lm(log(AAM)~AnnualTemp, data = sub)
  grass.linear.results[i,1]<-as.numeric(AIC(reg))
  grass.linear.results[i,2]<-summary(reg)$adj.r.squared
}


## Linear addictive models (same slope, different intercept)
grass.add.results <- matrix(NA,100,2)
for(i in 1:100){
  sub <- Grass.clean %>% group_by(Code) %>% sample_n(size=1)
  reg <- lm(log(AAM)~AnnualTemp+Condition, data = sub)
  grass.add.results[i,1]<-as.numeric(AIC(reg))
  grass.add.results[i,2]<-summary(reg)$adj.r.squared
}


## Interaction models (different slope, same intercept)
grass.int.results <- matrix(NA,100,2)
for(i in 1:100){
  sub <- Grass.clean %>% group_by(Code) %>% sample_n(size=1)
  reg <- lm(log(AAM)~AnnualTemp:Condition, data = sub)
  grass.int.results[i,1]<-as.numeric(AIC(reg))
  grass.int.results[i,2]<-summary(reg)$adj.r.squared
}


## Grouped-specific model (different slope, different intercept)
grass.group.results <- matrix(NA,100,2)
for(i in 1:100){
  sub <- Grass.clean %>% group_by(Code) %>% sample_n(size=1)
  reg <- lm(log(AAM)~AnnualTemp*Condition, data = sub)
  grass.group.results[i,1]<-as.numeric(AIC(reg))
  grass.group.results[i,2]<-summary(reg)$adj.r.squared
}

## Compare the AIC scores
mean(unique(grass.linear.results[,1]))
```

```
## [1] 9.652328
```

```r
mean(unique(grass.add.results[,1]))
```

```
## [1] 10.85402
```

```r
mean(unique(grass.int.results[,1]))
```

```
## [1] 11.02684
```

```r
mean(unique(grass.group.results[,1]))
```

```
## [1] 12.8044
```

```r
# R^2 value for the four models
r_2 <- data.frame(
  Model = c("Simple linear", "Linear additive", "Interaction", "Grouped"),
  R2 = c(mean(unique(grass.linear.results[,2])),
         mean(unique(grass.add.results[,2])),
         mean(unique(grass.int.results[,2])),
         mean(unique(grass.group.results[,2])))
)
kable(r_2)
```

Model	R2
Simple linear	0.6338739
Linear additive	0.6305711
Interaction	0.6278923
Grouped	0.6160965

- No preferences.

Using cold temperature

```r
## Simple linear model
grass.linear.results <- matrix(NA,100,2)
for(i in 1:100){
  sub <- Grass.clean %>% group_by(Code) %>% sample_n(size=1)
  reg <- lm(log(AAM)~ColdTemp, data = sub)
  grass.linear.results[i,1]<-as.numeric(AIC(reg))
  grass.linear.results[i,2]<-summary(reg)$adj.r.squared
}


## Linear addictive models (same slope, different intercept)
grass.add.results <- matrix(NA,100,2)
for(i in 1:100){
  sub <- Grass.clean %>% group_by(Code) %>% sample_n(size=1)
  reg <- lm(log(AAM)~ColdTemp+Condition, data = sub)
  grass.add.results[i,1]<-as.numeric(AIC(reg))
  grass.add.results[i,2]<-summary(reg)$adj.r.squared
}


## Interaction models (different slope, same intercept)
grass.int.results <- matrix(NA,100,2)
for(i in 1:100){
  sub <- Grass.clean %>% group_by(Code) %>% sample_n(size=1)
  reg <- lm(log(AAM)~ColdTemp:Condition, data = sub)
  grass.int.results[i,1]<-as.numeric(AIC(reg))
  grass.int.results[i,2]<-summary(reg)$adj.r.squared
}
```

```r
## Grouped-specific model (different slope, different intercept)
grass.group.results <- matrix(NA,100,2)
for(i in 1:100){
  sub <- Grass.clean %>% group_by(Code) %>% sample_n(size=1)
  reg <- lm(log(AAM)~ColdTemp*Condition, data = sub)
  grass.group.results[i,1]<-as.numeric(AIC(reg))
  grass.group.results[i,2]<-summary(reg)$adj.r.squared
}

## Compare the AIC scores
mean(unique(grass.linear.results[,1]))
```

```
## [1] 9.058041
```

```r
mean(unique(grass.add.results[,1]))
```

```
## [1] 11.02516
```

```r
mean(unique(grass.int.results[,1]))
```

```
## [1] 11.07396
```

```r
mean(unique(grass.group.results[,1]))
```

```
## [1] 12.89795
```

```r
# R^2 value for the four models
r_2 <- data.frame(
  Model = c("Simple linear", "Linear additive", "Interaction", "Grouped"),
  R2 = c(mean(unique(grass.linear.results[,2])),
         mean(unique(grass.add.results[,2])),
         mean(unique(grass.int.results[,2])),
         mean(unique(grass.group.results[,2])))
)
kable(r_2)
```

Model	R2
Simple linear	0.6419089
Linear additive	0.6283523
Interaction	0.6278365
Grouped	0.6149530

- No preference over the four models.
- While models using cold temperature does not show any significant differences compared to using annual temperature (AICs <= 2).

Bighead and silver carp

Using annual temperature

```r
Big.sil.clean <- Big.sil %>%
  filter(Condition %in% c("natural", "artificial"))

table(Big.sil.clean$Code)
```

```
##
##    A  AC  AD  AF  AG  AI  AL  AP   B   C   D   E   H   J   M   N   O   S   Z
##    4   1   1   2   2   1   1   1   3   1   2   2   1   1   2   1   2   2   2
```

```r
## Simple linear model
bs.linear.results <- matrix(NA,100,2)
for(i in 1:100){
  sub <- Big.sil.clean %>% group_by(Code) %>% sample_n(size=1)
  reg <- lm(log(AAM)~AnnualTemp, data = sub)
  bs.linear.results[i,1]<-as.numeric(AIC(reg))
  bs.linear.results[i,2]<-summary(reg)$adj.r.squared
}


## Linear addictive models (same slope, different intercept)
bs.add.results <- matrix(NA,100,2)
for(i in 1:100){
  sub <- Big.sil.clean %>% group_by(Code) %>% sample_n(size=1)
  reg <- lm(log(AAM)~AnnualTemp+Condition, data = sub)
  bs.add.results[i,1]<-as.numeric(AIC(reg))
  bs.add.results[i,2]<-summary(reg)$adj.r.squared
}


## Interaction models (different slope, same intercept)
bs.int.results <- matrix(NA,100,2)
for(i in 1:100){
  sub <- Big.sil.clean %>% group_by(Code) %>% sample_n(size=1)
  reg <- lm(log(AAM)~AnnualTemp:Condition, data = sub)
  bs.int.results[i,1]<-as.numeric(AIC(reg))
  bs.int.results[i,2]<-summary(reg)$adj.r.squared
}


## Grouped-specific model (different slope, different intercept)
bs.group.results <- matrix(NA,100,2)
for(i in 1:100){
  sub <- Big.sil.clean %>% group_by(Code) %>% sample_n(size=1)
  reg <- lm(log(AAM)~AnnualTemp*Condition, data = sub)
  bs.group.results[i,1]<-as.numeric(AIC(reg))
  bs.group.results[i,2]<-summary(reg)$adj.r.squared
}

## Compare the AIC scores
mean(unique(bs.linear.results[,1]))
```

```
## [1] 14.31244
```

```r
mean(unique(bs.add.results[,1]))
```

```
## [1] 12.44987
```

```r
mean(unique(bs.int.results[,1]))
```

```
## [1] 13.63488
```

```r
mean(unique(bs.group.results[,1]))
```

```
## [1] 14.08767
```

```r
# R^2 value for the four models
r_2 <- data.frame(
  Model = c("Simple linear", "Linear additive", "Interaction", "Grouped"),
  R2 = c(mean(unique(bs.linear.results[,2])),
         mean(unique(bs.add.results[,2])),
         mean(unique(bs.int.results[,2])),
         mean(unique(bs.group.results[,2])))
)
kable(r_2)
```

Model	R2
Simple linear	0.3367264
Linear additive	0.4217512
Interaction	0.3816686
Grouped	0.3899380

- For bighead and silver carp using annual air temperature, the simple linear model works the best. Although there does not seem to be much difference among the models.

Using cold temperature

```r
## Simple linear model
bs.linear.results <- matrix(NA,100,2)
for(i in 1:100){
  sub <- Big.sil.clean %>% group_by(Code) %>% sample_n(size=1)
  reg <- lm(log(AAM)~ColdTemp, data = sub)
  bs.linear.results[i,1]<-as.numeric(AIC(reg))
  bs.linear.results[i,2]<-summary(reg)$adj.r.squared
}


## Linear addictive models (same slope, different intercept)
bs.add.results <- matrix(NA,100,2)
for(i in 1:100){
  sub <- Big.sil.clean %>% group_by(Code) %>% sample_n(size=1)
  reg <- lm(log(AAM)~ColdTemp+Condition, data = sub)
  bs.add.results[i,1]<-as.numeric(AIC(reg))
  bs.add.results[i,2]<-summary(reg)$adj.r.squared
}


## Interaction models (different slope, same intercept)
bs.int.results <- matrix(NA,100,2)
for(i in 1:100){
  sub <- Big.sil.clean %>% group_by(Code) %>% sample_n(size=1)
  reg <- lm(log(AAM)~ColdTemp:Condition, data = sub)
  bs.int.results[i,1]<-as.numeric(AIC(reg))
  bs.int.results[i,2]<-summary(reg)$adj.r.squared
}
```

```r
## Grouped-specific model (different slope, different intercept)
bs.group.results <- matrix(NA,100,2)
for(i in 1:100){
  sub <- Big.sil.clean %>% group_by(Code) %>% sample_n(size=1)
  reg <- lm(log(AAM)~ColdTemp*Condition, data = sub)
  bs.group.results[i,1]<-as.numeric(AIC(reg))
  bs.group.results[i,2]<-summary(reg)$adj.r.squared
}

## Compare the AIC scores
mean(unique(bs.linear.results[,1]))
```

```
## [1] 17.12788
```

```r
mean(unique(bs.add.results[,1]))
```

```
## [1] 15.74328
```

```r
mean(unique(bs.int.results[,1]))
```

```
## [1] 18.90733
```

```r
mean(unique(bs.group.results[,1]))
```

```
## [1] 17.3182
```

```r
# R^2 value for the four models
r_2 <- data.frame(
  Model = c("Simple linear", "Linear additive", "Interaction", "Grouped"),
  R2 = c(mean(unique(bs.linear.results[,2])),
         mean(unique(bs.add.results[,2])),
         mean(unique(bs.int.results[,2])),
         mean(unique(bs.group.results[,2])))
)
kable(r_2)
```

Model	R2
Simple linear	0.2339017
Linear additive	0.3129891
Interaction	0.1892271
Grouped	0.2834452

- Same conclusion as annual temperature. Annual temperature fits better models (significant differences compared to using cold temperature).
- R^2 seems to be much lower.
- Linear additive model.
- Improved R^2 values.

Concluding points

- For all the other asian carp species, the simple linear model seems to be the best fit. Besides, using cold temperature and annual temperature does not produce different results.

- For black carp, the linear additive model (*same slope, different intercept*) seems to be a better fit. Using cold temperature on the original data suggests that the interaction model is the best (probably because the winter temperature at these two locations are very similar.)
- After removing the two data points, we see that the linear additive model still seems to be a better fit. Besides, greatly improved R^2 value suggests better fit.