North American Journal of Fisheries Management 36:1425–1436, 2016

DOI: 10.1080/02755947.2016.1214645

Supplement A: Example Statistical Code

Included here is details of the statistical code used to **(A)** calculate the biomass of adult and age-1+ Walleye in the lakes sampled; **(B)** calculate the differences in young-of-year (YOY) catch per unit effort (CPUE) during the pre- and post-2000 time periods; and **(C)** R-code for doing partial least square regression (PLSR) analysis. In this example, we look at results from Lake Vermilion. All of the code here was used in R version 3.1.0 (“Spring Dance”), through the RStudio interface (version 0.98.953).

(A) *Calculating the biomass of adult and age-1+ Walleyes.* Our method of calculating Walleye biomass uses the models previously described by Anderson (1998). In R code, any text following the pound sign (#) is not part of the code and instead represents a note to the reader. Detailed notes are included for the first example of processes that are repeated for each lake.

```
##QAbG data generation
##QAbG data generation
##QAbG data generation
library("plyr")  #This package allows for easier data manipulation.

MDNR_all_fish <- read.csv("~/USGS Science/VOYA
Fish/Analysis/MDNR_all_fish.csv")

#A database compiled from MDNR large lakes

#Here we create a set of unique identifiers for every recorded seine
#haul or gill net
Isolatingdata <-data.frame(MDNR_all_fish$"START_DATE")
Isolatingdata$Station<-MDNR_all_fish$SAMP_STA_FULL_NAME
Isolatingdata$Lake<-MDNR_all_fish$LAKE_NAME
Isolatingdata$Type<-MDNR_all_fish$EFFORT_UNITS
Isolatingdata$Rep<-MDNR_all_fish$REPLICATE
Isolatingdata$Effort<-MDNR_all_fish$EFFORT
#The “Isolatingdata” data frame is just a combination of all information
#for each effort (gill net or seine haul)
UniqueIDs <-apply(Isolatingdata,1,paste,collapse=" ")
UniqueIDs <-data.frame(UniqueIDs)
MDNRAllUniques <-cbind(MDNR_all_fish,UniqueIDs)
#And here we create a new data frame with each effort uniquely tagged

MDNRAllUniques <- subset(MDNRAllUniques, UniqueIDs != "1996-07-10 SSE-3 Lake
of the Woods Haul  1  1" )
#This particular seine haul is clearly incorrect (>100,000 individuals of
each
```

```
# species recorded) and is therefore removed.
MDNRAllUniques <- subset(MDNRAllUniques, LAKE_NAME != "Cass" )
MDNRAllUniques <- subset(MDNRAllUniques, LAKE_NAME != "Crane" )
MDNRAllUniques <- subset(MDNRAllUniques, LAKE_NAME != "Namakan" )
MDNRAllUniques <- subset(MDNRAllUniques, LAKE_NAME != "Sand Point" )
#These lakes don't have enough data to be useful

## Starting with Winnibigoshish##
## Starting with Winnibigoshish##
## Starting with Winnibigoshish##
## Starting with Winnibigoshish##

# For qabg & CPUE calculations from D. Staples/MNDNR
#   need Individual fish records with
#   MM, GRAM, YEAR, NET & SEX
#==============================================

WinnibigoshishsetdataWall <- subset(MDNRAllUniques,YEAR >'1982' &
LAKE_NAME=='Winnibigoshish' & EFFORT_UNITS=='Set' &
FISH_SPECIES_ABBREV=='WAE')
WinnibigoshishsetdataWall <- subset(WinnibigoshishsetdataWall,WEIGHT_G>0)

WinnibigoshishData <- WinnibigoshishsetdataWall$YEAR
WinnibigoshishData <-data.frame(WinnibigoshishData)
colnames(WinnibigoshishData)<- "YEAR"
WinnibigoshishData$MM<-WinnibigoshishsetdataWall$TOTAL_LENGTH_MM
WinnibigoshishData$GRAM<-WinnibigoshishsetdataWall$WEIGHT_G
WinnibigoshishData$NET<-WinnibigoshishsetdataWall$UniqueIDs
WinnibigoshishData$SEX<-WinnibigoshishsetdataWall$GENDER_TYPE_ABBREV

#==============================================
#     Input Data
LakeName = 'Winnibigoshish'
PSL.year = 2000
acres = 56470  # lake size

####  vector of number of nets set in each year
num.nets = c(rep(24,8),22,rep(24,3),    # 1983-1994
             23,rep(24,2),23,            # to 1998
             23,rep(24,3),23,rep(24,2),  # to 2005
             rep(22,2),24,24)         # to 2009

###   GILLNET DATA
# For qabg & CPUE calculations,
#   need Individual fish records with
#   MM, GRAM, YEAR, NET & SEX
Data = WinnibigoshishData
#==============================================

num.yrs = length(num.nets)
start.yr = min(Data$YEAR)
end.yr = max(Data$YEAR)
Data$Yr = as.factor(Data$YEAR)
Data$PSL[Data$YEAR < PSL.year] = 0
Data$PSL[Data$YEAR >= PSL.year] = 1
Data$PSL = as.factor(Data$PSL)
```

```
Data$SEX[Data$SEX == '2'] = 'M' #Conversion based on comparison between this
data and data provided by D. Staples for Lake Winnibigoshish
Data$SEX[Data$SEX == '1'] = 'F'

sums = apply(table(Data$YEAR,Data$NET),1,sum)
cpue = sums/num.nets

#  Males and females were not separated here.
length(Data$GRAM)-length(na.omit(Data$GRAM)) #12 missing Gram data in data,
just removing them
Data<-Data[complete.cases(Data),]

#=======================================
#  Calculate qabg adjustments for
#   range of fish lengths (200-799 mm by 1 mm)
#======================================
# Gill-net selectivity functions written by
# C. S. Anderson and based on: Qabg
# Anderson (1998).
# 21 September 2009 revised CSA so that
#    coefficients match the "revised 1999"
#    values as in Dale Logsdon's excelqabg.xls
# Key results:
# tl = a vector of tl values, now 200:799
# qabg.total = a vector of qabg (all three
#      parts) * 10^5  (good for Mille Lacs)
# qbg.total = a vector of qbg (no approach part)
# qbgcor.total = same as qbg.total except
#      the values have been multiplied by
#          correction factor for 575 < tl < 800.

# Specify coefficients
# mesh bar measure
mesh <- c(19,25,32,38,51)

# total lengths
tl <- 200:799     # Can change TL range here

###  contact coefficients beta[]
winni.b <- c(0.059,0.130,0.255,0.472,1.0)
#ml.b <- c(0.173, 0.377, 0.717, 0.867, 1.0) # revised 1999
#oldml.b <- c(0.185,0.392,0.734,0.896,1.0)
#low.b <- c(0.311,0.703,0.855,0.921,1.0)
#rainy.b <- c(0.396,0.580,0.601,0.656,1.0)
beta <- winni.b   # Can change lake here.

### retention coefficients b[]
winni <- c(-3.467,-0.988,1.804,0.289,24.450)
#ml <- c(-2.166, -1.024, 1.123, 0.288, 14.48451) # revised 1999
#oldml <- c(-2.053,-1.013,1.058,0.280,15.209)
#low <- c(-1.880,-0.976,1.008,0.261,28.807)
#rainy <- c(-2.119,-0.897,1.214,0.235,57.813)
b <-winni   # Can change lake here.

# Specify functions
app <- function(tl) {
  a <- 0.3535
```

```
  b <- (345.4 < tl) * 0.008695 * (tl-345.4)
  correction <- (575 < tl & tl < 800) *( -(tl-575)/(800-575) )
  (a+b)*(1+correction)
}

# retention function
ret <- function(b,x) {
  exp((b[1]+b[3]*x)/(1+b[2]*x+b[4]*x^2))/b[5]
}

# retention function with correction at large TL
retcorr <- function(b,x) {
  a <- exp((b[1]+b[3]*x)/(1+b[2]*x+b[4]*x^2))/b[5]
  correction <- (575 < tl & tl < 800) *( -(tl-575)/(800-575) )
  a*(1+correction)
}

# Calculate x = TL/mesh perimeter ratio
x <- outer(tl,mesh,"/") / 4.0

# Selectivity components
alpha <- app(tl)  # approach prob. *10^5
retent <- ret(b,x)      # retention curve.
retentcor <- retcorr(b,x)     # retention curve with correction
qbg <- sweep(retent,2,beta,"*")     # indirect sel.
qbg.total <- apply(qbg,1,sum)
qbgcor <- sweep(retentcor,2,beta,"*")     # indirect sel. with correction
qbgcor.total <- apply(qbgcor,1,sum)
qabg <- sweep(qbg,1,alpha,"*")      # Qabg * 10^5
qabg.total <- apply(qabg,1,sum)
#  Above code outputs vectors of qabg and qbg(corr) that
#  are used below to estimate abundance by:
#     Nhat = CPE/qabg.total * Acres/132516 * 10^5
#  then size-specific estimates are summed
#  for estimate of overall total


#========================================
#  Calculate qabg-adjusted estimates
#========================================
##  boths
LY.both = with(Data[Data$MM>199,], table(MM,Yr))
tl.f = as.numeric(rownames(LY.both))
ndx.f = match(tl.f, tl)
qabg.total.f = qabg.total[ndx.f]
qbg.total.f = qbg.total[ndx.f]
qbgcor.total.f = qbgcor.total[ndx.f]
KG.both.m = with(Data[Data$MM>199,], lm(log(GRAM) ~ I(log(MM))))
KG.both = exp(KG.both.m$coef[1] + log(tl.f)*KG.both.m$coef[2])
# Output variables
f.qabg.size = matrix(0,length(ndx.f),num.yrs)
rownames(f.qabg.size) = tl.f
colnames(f.qabg.size) = levels(Data$Yr)
f.qabgKG.size = matrix(0,length(ndx.f),num.yrs)
rownames(f.qabgKG.size) = tl.f
colnames(f.qabgKG.size) = levels(Data$Yr)
f.qabg = rep(NA,num.yrs)
```

```
f.qabgKG = rep(NA,num.yrs)
f.qbg = rep(NA,num.yrs)
f.qbgKG = rep(NA,num.yrs)
f.qbgcor = rep(NA,num.yrs)
f.qbgcorKG = rep(NA,num.yrs)
f.ssb = rep(NA,num.yrs)
mature.size = which(tl.f > 406)[1]
# loop through years and calculate statistics
for(i in 1:num.yrs)
{
  cpe = LY.both[,i]/num.nets[i]
  #  qabg
  qabg = cpe / qabg.total.f * acres/132516 * 10^5
  f.qabg[i] = sum(qabg,na.rm=T)
  f.qabg.size[,i] = qabg
  qabgKG = qabg * KG.both
  f.qabgKG[i] = sum(qabgKG,na.rm=T)
  f.ssb[i] = sum(qabgKG[mature.size:end(tl.f)[1]],na.rm=T)
  f.qabgKG.size[,i] = qabgKG
  #  qbg
  qbg = cpe / qbg.total.f * acres/132516 * 10^5
  f.qbg[i] = sum(qbg,na.rm=T)
  qbgKG = qbg * KG.both
  f.qbgKG[i] = sum(qbgKG,na.rm=T)
  #  qbgcor
  qbgcor = cpe / qbgcor.total.f * acres/132516 * 10^5
  f.qbgcor[i] = sum(qbgcor,na.rm=T)
  qbgcorKG = qbgcor * KG.both
  f.qbgcorKG[i] = sum(qbgcorKG,na.rm=T)
}
f.qabg.size = round(f.qabg.size,1)

#  create & save data frame w/ estimates
Winnibigoshish.qabg = as.data.frame(list(
  YEAR = c(start.yr:end.yr),
  f.ssb = f.ssb,
  f.qabg = f.qabg,
  f.qabgKG = f.qabgKG,
  f.qbg = f.qbg,
  f.qbgKG = f.qbgKG,
  f.qbgcor = f.qbgcor,
  f.qbgcorKG = f.qbgcorKG,
  gillnet.cpue = cpue
))
Winnibigoshish.qabg$Total<-Winnibigoshish.qabg$f.qabgKG
##Biomass of all Walleye
Winnibigoshish.qabg$TotalAdult<-Winnibigoshish.qabg$f.ssb
##Biomass of adults
#  save(Winnibigoshish.qabg,file='Winnibigoshish.qabg')
#  load('WinnieQABG.RData')

##This repeats for all of the lakes individually

rownames(Vermilion.qabg) <- NULL
rownames(Winnibigoshish.qabg) <- NULL
rownames(Rainy.qabg) <- NULL
rownames(Kabetogama.qabg) <- NULL
```

```
rownames(Leech.qabg) <- NULL
rownames(LakeoftheWoods.qabg) <- NULL

##Then we create a combined file that includes all data from all of
## the lakes

Total<-
rbind(Vermilion.qabg,Winnibigoshish.qabg,Rainy.qabg,Kabetogama.qabg,Leech.qab
g,LakeoftheWoods.qabg)
write.table(Total,file="~/USGS Science/VOYA
Fish/Analysis/Walleyeqabg.csv",sep=",")
```

(B) *Calculating the differences in age-0 CPUE between the pre- and post-2000 time periods.* In this example, we estimate the difference in age-0 Walleye CPUEs calculated for Lake Kabetogama during the pre- and post-2000 periods. The same procedure was performed for each lake and for the age-0 Yellow Perch CPUE data.

```
library("plyr")
library("BEST") #This package performs the BEST analysis

AllWalleyeYOYAdult<-read.csv("~/USGS Science/VOYA
Fish/Analysis/YOYandAdultWalleye.csv")
##We shortened the lake names for the purpose of visualizing later
AllWalleyeYOYAdult$Lake<-mapvalues(AllWalleyeYOYAdult$Lake, from =
c("Kabetogama","LakeoftheWoods","Leech","Rainy","Vermilion","Winnibigoshish")
, to = c("K","LOW","L","R","V","W"))
## there are some infinite values that should be NA
is.na(AllWalleyeYOYAdult) <- do.call(cbind,lapply(AllWalleyeYOYAdult,
is.infinite))
## there are some zero values that should be NA
AllWalleyeYOYAdult$RperS<-
AllWalleyeYOYAdult$YOYCPUE/AllWalleyeYOYAdult$AdultCPUE

Walleye1990s<- subset(AllWalleyeYOYAdult,YEAR<2000)
Walleye1990s<- subset(Walleye1990s,YEAR>1989)

Walleye1990sAverages <-
ddply(Walleye1990s,.(Lake),summarize,Average=mean(YOYCPUE,na.rm=TRUE))
Walleye1990sStDev <-
ddply(Walleye1990s,.(Lake),summarize,StDev=sd(YOYCPUE,na.rm=TRUE))
Walleye1990sAverages<- merge(Walleye1990sAverages,Walleye1990sStDev, by =
"Lake")

Walleye2000s<- subset(AllWalleyeYOYAdult,YEAR>2000)

Walleye2000sAverages <-
ddply(Walleye2000s,.(Lake),summarize,Average=mean(YOYCPUE,na.rm=TRUE))
Walleye2000sStDev <-
ddply(Walleye2000s,.(Lake),summarize,StDev=sd(YOYCPUE,na.rm=TRUE))
Walleye2000sAverages<- merge(Walleye2000sAverages,Walleye2000sStDev, by =
"Lake")

Kabetogama1990s<- subset(AllWalleyeYOYAdult,YEAR>1989 & Lake == "K")
```

```
Kabetogama1990s<- subset(Kabetogama1990s,YEAR<2000)
Kabetogama1990s$Era <- "1990s"
Kabetogama2000s<- subset(AllWalleyeYOYAdult,YEAR>1999 & Lake == "K")
Kabetogama2000s$Era <- "2000s"
KabetogamaPrePost<-rbind(Kabetogama1990s,Kabetogama2000s)

Kabetogama2000s<-Kabetogama2000s[complete.cases(Kabetogama2000s),]
#getting rid of NAs that seem to screw up the analysis

BESTout <- BESTmcmc(Kabetogama1990s$YOYCPUE, Kabetogama2000s$YOYCPUE)
BESTout
summary(BESTout)
plot(BESTout)
```

(C) *Conducting partial least-squares regression analysis with R code*. In this example, we perform partial least-squares regression for the Lake Vermilion data and examine the results.

```
library("plyr")
library("pls")
#Packages used in this code

CPUEandWLData<-read.csv("~/USGS Science/VOYA
Fish/Analysis/WalleyeCPUEandWLData.csv")

#These are the data as included in the Data Supplement spreadsheet
#“WalleyeCPUEandWLData”

CPUEandWLData$lnAdultCPUE<- log(CPUEandWLData$AdultCPUE)
CPUEandWLData$lnYOYCPUE<- log(CPUEandWLData$YOYCPUE)
CPUEandWLData$lnAdultqabg <- log(CPUEandWLData$TotalAdult)

#Log-transforming the response variable

#Starting with Lake Vermilion

Vermilion<-subset(CPUEandWLData,CPUEandWLData$Lake == "Vermilion")


VermilionNumeric<- Vermilion
# Creating a file that does not include non-numeric data

VermilionNumeric$Lake <- NULL
VermilionNumeric$GNEffort <- NULL


RStd <- scale(VermilionNumeric)

#Scaling the data like this cannot be done with non-numeric data

Vermilionstd <- data.frame(RStd)

colMeans(Vermilionstd)

#Check to make sure that the means are essentially zero
```

```
ColumnHeadings<-colnames(Vermilionstd)

#Obtaining a list of column names so that we can specify the models more
easily

VermilionWalleyeYOY.PLS<- plsr(lnYOYCPUE ~
AdultCPUE+Total1PlusCPUE+MinWL+WLR+MeanWL+

StDevWL+PrevSummerElevation+SpringRise+SpringMean+IceoutElevation+DDZero

+DD5+DD10+DD15+SprDDZero+SprDD5+SprDD10+SprDD15+TimeFromIO+WinterDecline

+f.qabgKG+SpringSD+SpringCV+SpringCVWL+SpringSDWL,
             data = Vermilionstd,
             validation = "CV",scale=FALSE,na.action = na.exclude)

#Specifying the model.  Scale is false because we already scaled the data.
#uses the standard cross-validation method “CV”

summary(VermilionWalleyeYOY.PLS)

#RMSEP suggests that 10 components could be included,
#just two components explain ~92.6% of the variation, three explain 94.6%,
#others add little explanatory power (<3%)

VermilionWalleyeYOY.PLSLoadings<-
round(as.data.frame(unclass(loadings(VermilionWalleyeYOY.PLS))),2)

#Save the loadings as a data frame file for export
```

SUPPLEMENTAL REFERENCES

Anderson, C. S. 1998. Partitioning total size selectivity of gill nets for Walleye (*Stizostedion vitreum*) into encounter, contact, and retention components. Canadian Journal of Fisheries and Aquatic Sciences 55:1854–1863.