Zooplankton Data Input for LCM

Brian Mahardja

2023-04-12

Original script and dataset packages was written by Sam Bashevkin and can be found here: <https://github.com/sbashevkin/FLOATDrought>

# Load packages

require(conflicted)  
require(MASS)  
require(dplyr)  
require(zooper)  
require(lubridate)  
require(readr)  
require(tidyr)  
require(ggplot2)  
require(sf)  
require(readxl)  
require(stringr)  
require(mgcv)  
require(purrr)  
require(deltamapr)  
require(scales)  
  
conflict\_prefer("filter", "dplyr")  
conflict\_prefer("select", "dplyr")

# Load data

Load zooplankton model output and strata volume from Will Smith

model\_data<-read.csv("zoop sal adjustments.csv")

# Adjust data per Compass Memo

When 0s were introduced for baseline predictions, it resulted in infinite values for the scalars. We replaced these infinities with the maximum finite scalar calculated from model predictions for a specific action, taxon, subregion, and month (across years). When this step still yielded no finite scalar value for a given action, taxon, subregion, and month, we used the maximum finite scalar value from +/- one month.

##Get the maximum based on region and month  
  
#Change infinite to NA to remove it from max calculation  
model\_data\_maximum\_region\_month <- model\_data  
model\_data\_maximum\_region\_month[model\_data\_maximum\_region\_month == Inf] <- NA  
model\_data\_maximum\_region\_month <- na.omit(model\_data\_maximum\_region\_month)   
  
model\_data\_maximum\_region\_month <- model\_data\_maximum\_region\_month%>% select(-year) %>% group\_by(region,month,IBMR) %>% summarise\_all(max) %>%  
 pivot\_longer(cols=starts\_with("sal\_"), names\_to = "variable", values\_to = "replaceValue\_Region\_Month")  
  
  
##Get the maximum based on month  
#Remove NA and change infinite to 0 to remove it from max calculation  
model\_data\_maximum\_month <- model\_data  
model\_data\_maximum\_month[model\_data\_maximum\_month == Inf] <- NA  
model\_data\_maximum\_month <- na.omit(model\_data\_maximum\_month)   
  
model\_data\_maximum\_month <- model\_data\_maximum\_month %>% select(-year,-region) %>% group\_by(month,IBMR) %>% summarise\_all(max) %>%  
 pivot\_longer(cols=starts\_with("sal\_"), names\_to = "variable", values\_to = "replaceValue\_Month")  
  
##Adjust data to allow for imputation  
model\_data\_adjusted<- model\_data %>%  
 pivot\_longer(cols=starts\_with("sal\_"), names\_to = "variable", values\_to = "Value")  
#Join data  
model\_data\_adjusted <- model\_data\_adjusted %>% left\_join(model\_data\_maximum\_region\_month) %>% left\_join(model\_data\_maximum\_month)  
##Fill in data  
model\_data\_adjusted$AdjustedValue <- ifelse(model\_data\_adjusted$Value==Inf,ifelse(is.na(model\_data\_adjusted$replaceValue\_Region\_Month),model\_data\_adjusted$replaceValue\_Month,model\_data\_adjusted$replaceValue\_Region\_Month),model\_data\_adjusted$Value)  
  
#Remove extra columns and pivot wider  
model\_data\_adjusted <- model\_data\_adjusted %>% select(-c(replaceValue\_Month,replaceValue\_Region\_Month,Value)) %>% rename(scenario = variable)

# Convert IBMR region to LCM region

#Information from Will Smith (USFWS)  
# spatial strata names: Yolo Sac SDelta EDelta LowSac LowSJ Conf SSuisunE NSuisunE Marsh SSuisunW NSuisunW  
#vol<-c(100510941,144940444,225381539,65449147,89444941,259500691,163419100,153797952,76796487,121672916,107178813,184941122) # DSM2 volumes  
  
  
volume\_IBMR<-data.frame(region=c("Confluence","Suisun Marsh","NE Suisun","SE Suisun","NW Suisun","SW Suisun"),volume=c(163419100,121672916,76796487,153797952,184941122,107178813),LCM\_region=c("West","West","West","West","West","Far West"))  
  
#SW Suisun will be used to represent the Far West, everything else will represent West region  
  
#Divide West region by proportion  
volume\_IBMR$vol\_proportion<-ifelse(volume\_IBMR$LCM\_region=="West",volume\_IBMR$volume/(163419100+121672916+76796487+153797952+184941122),1)  
  
#Join data  
model\_data\_adjusted <- model\_data\_adjusted %>% left\_join(volume\_IBMR)  
  
#Break down by LCM region  
model\_data\_adjusted\_LCM <- model\_data\_adjusted %>% mutate(AdjustedValue\_prop = AdjustedValue\*vol\_proportion) %>% group\_by(year, month, LCM\_region, scenario, IBMR) %>%   
 summarise(AdjustedValue = sum(AdjustedValue\_prop))

# Use LCM zooplankton data input to calculate proportion of taxa

original\_zoop\_data <-read.csv("../Zooplankton\_input\_original/ZooMysid\_74\_19\_df.csv")  
   
zoop\_data\_IBMR\_config <- original\_zoop\_data  
zoop\_data\_IBMR\_config[is.na(zoop\_data\_IBMR\_config)] <- 0  
  
zoop\_data\_IBMR\_config <- zoop\_data\_IBMR\_config %>% mutate(acartela = ACARTELA,  
 allcopnaup = COPNAUP+OTHCOPNAUP+EURYNAUP+SINONAUP+PDIAPNAUP,  
 daphnia = DAPHNIA,  
 eurytem = EURYTEM,  
 limno = LIMNOSPP+LIMNOSINE+LIMNOTET,  
 mysid = Hyperacanthomysis.longirostris+Neomysis.mercedis,  
 othcalad = OTHCALAD+SINOCAL,  
 othcaljuv = CALJUV+OTHCALJUV+EURYJUV+SINOCALJUV+PDIAPJUV+ASINEJUV+ACARJUV+DIAPTJUV+TORTJUV,  
 othclad=BOSMINA+DIAPHAN+OTHCLADO,  
 othcyc = AVERNAL,  
 pdiapfor = PDIAPFOR  
 ) %>%   
 select(Year,Date,Region,acartela,allcopnaup,daphnia,eurytem,limno,mysid,othcalad,othcaljuv,othclad,othcyc,pdiapfor) %>%  
 mutate(Month=month(Date)) %>% select(-Date) %>%  
 group\_by(Year,Month,Region) %>% summarise\_all(list(median)) %>%  
 #calculate proportion of taxa from each sample  
 mutate(total=acartela+allcopnaup+daphnia+eurytem+limno+mysid+othcalad+othcaljuv+othclad+othcyc+pdiapfor) %>%  
 mutate\_at(vars(acartela:pdiapfor),list(~./total)) %>% select(-total) %>%  
 pivot\_longer(cols=c("acartela","allcopnaup","daphnia","eurytem","limno","mysid","othcalad","othcaljuv","othclad","othcyc","pdiapfor"),  
 names\_to="IBMR", values\_to="zoop\_taxa\_ratio") %>%  
 rename(year=Year, month=Month, LCM\_region=Region)  
  
  
#Missing data to be filled with average 1995-2014 ratio for each region  
zoop\_data\_IBMR\_config\_region\_mean <- zoop\_data\_IBMR\_config %>% filter(year %in% c(1995:2014)) %>% group\_by(LCM\_region, IBMR) %>%   
 summarise(zoop\_taxa\_ratio\_mean = mean(zoop\_taxa\_ratio,na.rm=T))  
  
model\_data\_for\_LCM <- model\_data\_adjusted\_LCM %>% left\_join(zoop\_data\_IBMR\_config) %>% left\_join(zoop\_data\_IBMR\_config\_region\_mean)  
  
model\_data\_for\_LCM$zoop\_taxa\_ratio = ifelse(is.na(model\_data\_for\_LCM$zoop\_taxa\_ratio),model\_data\_for\_LCM$zoop\_taxa\_ratio\_mean,model\_data\_for\_LCM$zoop\_taxa\_ratio)  
  
#Other taxa from IBMR does not get used in the LCM, so we can safely remove it  
model\_data\_for\_LCM <- model\_data\_for\_LCM %>% select(-zoop\_taxa\_ratio\_mean) %>% filter(IBMR!="other")  
  
#Areas and taxa without predictions are left as is (Adjust Value of 1)  
model\_data\_for\_LCM$AdjustedValue<- ifelse(is.na(model\_data\_for\_LCM$AdjustedValue),1,model\_data\_for\_LCM$AdjustedValue)  
  
#Create the multiplier  
model\_data\_for\_LCM$Multiplier <- model\_data\_for\_LCM$AdjustedValue\*model\_data\_for\_LCM$zoop\_taxa\_ratio  
  
#Pivot wider to calculate the LCM taxa groups  
unique(model\_data\_for\_LCM$IBMR)  
  
#Adjust model\_data\_for\_LCM\_adjusted to allow for easy conversion  
model\_data\_for\_LCM\_adjusted <- model\_data\_for\_LCM %>% select(-zoop\_taxa\_ratio,-AdjustedValue) %>%  
 pivot\_wider(names\_from=IBMR,values\_from=Multiplier) %>% mutate(  
 NJ\_BPUV = allcopnaup + othcaljuv,  
 JA\_BPUV= acartela + eurytem + limno + othcalad + othcaljuv +othcyc + pdiapfor,  
 JAC\_BPUV= acartela + daphnia + eurytem + limno + othcalad + othcaljuv + othclad +othcyc + pdiapfor,  
 NJAC\_BPUV = acartela + allcopnaup + daphnia + eurytem + limno + othcalad + othcaljuv + othclad +othcyc + pdiapfor,  
 M\_BPUV = mysid,  
 JACM\_BPUV = acartela + daphnia + eurytem + limno + mysid + othcalad + othcaljuv + othclad +othcyc + pdiapfor,  
 NJACM\_BPUV =acartela + allcopnaup + daphnia + eurytem + limno + mysid + othcalad + othcaljuv + othclad +othcyc + pdiapfor,  
 ACM\_BPUV = acartela + daphnia + eurytem + limno + mysid + othcalad + othcaljuv + othclad +othcyc + pdiapfor,  
 ) %>%  
 select(-acartela,-allcopnaup,-daphnia,-eurytem,-limno,-mysid,-othcalad,-othcaljuv,-othclad,-othcyc,-pdiapfor) %>%  
 pivot\_longer(cols=ends\_with("BPUV"), names\_to = "Taxa", values\_to = "multiplier") %>%   
 #Rename columns  
 rename(Year=year,Month=month,Region=LCM\_region) %>% mutate(Month = as.character(month(Month, label = TRUE, abbr = FALSE)))

# Apply results according to LCM zooplankton dataset configurations

#Read data from LCM  
zoop\_data\_LCM\_raw <-read.csv("../Zooplankton\_input\_original/ZooMysid\_74\_19\_df\_median.csv")   
zoop\_data\_LCM <- zoop\_data\_LCM\_raw %>% filter(Year %in% c(1995:2017))  
  
#Find maximum value for each data  
zoop\_data\_LCM\_max <- zoop\_data\_LCM\_raw %>% pivot\_longer(cols=ends\_with("BPUV"), names\_to = "Taxa", values\_to = "BPUV") %>%  
 group\_by(Taxa) %>% summarise(BPUV\_max=max(BPUV,na.rm=T))  
  
#Create function to expand data frame based on a column  
expand.grid.df <- function(...) Reduce(function(...) merge(..., by=NULL), list(...))  
  
#Create data frame for conversion  
zoop\_data\_LCM\_expanded <- expand.grid.df(zoop\_data\_LCM,unique(model\_data\_for\_LCM\_adjusted$scenario)) %>% rename(scenario=y) %>%  
 pivot\_longer(cols=ends\_with("BPUV"), names\_to = "Taxa", values\_to = "BPUV") %>%  
 left\_join(model\_data\_for\_LCM\_adjusted)

## Joining with `by = join\_by(Year, Month, Region, scenario, Taxa)`

#Multiply values by multiplier. If N/A, it's because they're in the North and South regions, which should remain unaffected.  
zoop\_data\_LCM\_scenarios <- zoop\_data\_LCM\_expanded %>% mutate(BPUV\_adjusted= ifelse(is.na(multiplier),BPUV,BPUV\*multiplier)) %>%  
 #Adjust all values to within the bounds of maximum amount seen  
 left\_join(zoop\_data\_LCM\_max) %>% mutate(BPUV\_adjusted=ifelse(BPUV\_adjusted>BPUV\_max, BPUV\_max, BPUV\_adjusted)) %>%  
 select(-BPUV,-BPUV\_max,-multiplier) %>%  
 #Reformat data to original setup  
 pivot\_wider(names\_from=Taxa,values\_from=BPUV\_adjusted) %>%  
 #Remove baseline  
 filter(!(scenario %in% c("sal\_base\_median","sal\_base\_l95","sal\_base\_u95")))

## Joining with `by = join\_by(Taxa)`

#Export the full dataset  
write.csv(zoop\_data\_LCM\_scenarios, file="ZooMysid\_2022ROC\_EffectsAnalysis\_df\_median.csv", row.names=F)  
  
#Further aggregate and subset data per email with Will Smith (USFWS)  
#Use cohort year (time is indexed by cohort year, with the first month of the year beginning in April)  
  
  
zoop\_data\_feb\_mar\_ACM <- zoop\_data\_LCM\_scenarios %>% mutate(Cohort\_Year=ifelse(Month %in% c("January","February","March"),Year-1,Year)) %>%  
 group\_by(Cohort\_Year,Year,Month,scenario) %>% summarise(ACM\_BPUV=mean(ACM\_BPUV)) %>%  
 filter(Month %in% c("February","March")) %>% pivot\_wider(names\_from = Month, values\_from = ACM\_BPUV) %>%  
 rename(ACM\_BPUV\_Feb1Feb1 = February, ACM\_BPUV\_Mar1Mar1 = March,Calendar\_Year = Year) %>%  
 mutate(scenario=gsub("sal\_","",scenario)) %>%  
 mutate(zoop\_model\_prediction= case\_when(grepl("l95",scenario) ~ "l95",  
 grepl("median",scenario) ~ "median",  
 grepl("u95",scenario) ~ "u95")) %>%  
 mutate(scenario=gsub("\_l95","",scenario)) %>%  
 mutate(scenario=gsub("\_median","",scenario)) %>%  
 mutate(scenario=gsub("\_u95","",scenario)) %>% filter(Cohort\_Year>1994)

## `summarise()` has grouped output by 'Cohort\_Year', 'Year', 'Month'. You can  
## override using the `.groups` argument.

#Export the final zoop dataset  
write.csv(zoop\_data\_feb\_mar\_ACM, file="ZooMysid\_2022ROC\_EffectsAnalysis\_Feb\_Mar\_ACM.csv", row.names=F)  
  
#Combine with flow dataset  
zoop\_flow\_data\_LCM <- zoop\_data\_feb\_mar\_ACM %>% ungroup() %>% filter(zoop\_model\_prediction=="median") %>%  
 select(Cohort\_Year,scenario,ACM\_BPUV\_Feb1Feb1,ACM\_BPUV\_Mar1Mar1) %>% filter(Cohort\_Year>1994) %>%  
 left\_join(read.csv("../Hydro model output/FlowData\_2022ROC\_EffectsAnalysis\_CohortYear.csv"))

## Joining with `by = join\_by(Cohort\_Year, scenario)`

#Export the final flow and zoop dataset  
write.csv(zoop\_flow\_data\_LCM, file="FlowZoopData\_2022ROC\_EffectsAnalysis\_CohortYear.csv", row.names=F)