Cowlitz SAR Analysis

Thomas Buehrens & Co-authors (TBD)

Contents

This document was generated on 12/16/2020.

Purpose

The purpose of this document outline use of functions in this repository to estimate smolt to adult return rates (SAR) of salmon and steelhead using coded wire tag data located in RMIS **(link)**. Due to a very limited API in RMIS, this repository adopt a workflow where initial functions are used to pull a mirror of RMIS data to a local (or server) location designated by the user before querying it for subsequent summarization or analysis.

Requirements

All analyses require R software **(link)** (v3.4.3) for data retrieval, data processing, and summarizing model results. JAGs software **(link)** (v4.2.0) is used for fitting “group” based SAR models using for Markov chain Monte Carlo (MCMC) simulation), but is not necessary for analyses based on individual tag codes.

Functions

We also need a couple of helper functions which we will load from the functions folder, which we will load using the walk() function from the purrr package (which we will install if it is not already installed).

```r
#==========================================
# Source function files in functions folder
#==========================================
if(!require("purrr")) {
  install.packages("purrr")
  library("purrr")
}else(library(purrr))
path <- "functions/"
files <- paste0(path, list.files(path))
purrr::walk(files, source)
```

Packages

In addition to purr, We also need a few packages that are not included with the base installation of R, so we begin by installing them (if necessary) and then loading them.

```r
#===============================================
# Load packages, install and load if not already
#===============================================
using("lubridate", "openxlsx","tidyverse","readxl","scales","knitr")
if("MASS" %in% (.packages())){
  detach("package:MASS", unload=TRUE)
}
```

Dowloading And Initial Filter of Data

In this section we will use functions to complete an initial filter of CWT data for our SAR analysis. 1. If RMIS data has already been downloaded and is located in the “GatherData” folder, then “download_releases” and “download_recoveries” can be set to “No” and filtering can proceed without downloading data. If data has not yet been downloaded, these arguments should be set to “Yes”. 2. Users should then specify: a) the first and last brood years of interest, b) the minimum and maximum total fish age salmon or steelhead from they would like to look for recoveries for, c) the hatchery name (does not have to be exact), d) the CWT mark type (CWT=“CWT”), e) the clip type (“Ad” = ad clipped; query currently doesn’t support un-clipped fish). 3. The user should designate a filename for the release and recovery data to be written to 4. Optionally, a user can specify a run timing, multiple comma separated run timings in vector form, or leave run timing blank. The function will optionally download RMIS data to your “GatherData” directory (which it will create on the fly if it doesn’t already exist), query the data, and write the queried subset of raw RMIS data to an excel file in the “results” directory, which it will create on the fly if it doesn’t exist.

```
getdata(download_releases = "No",
        download_recoveries = "No",
        first_brood_year = 1980,
        last_brood_year = 2014,
        min_total_age = 3,
        max_total_age = 6,
        hatcheryname =c("COWLITZ"), #can use a comma separated list here
        SP = "Chinook",
        CWT = "CWT",
        marktype = "Ad",
        filename ="CowlitzSpChk_CWT_data",
        #optional argument (leave out if you don't want to limit)
        runtiming =c("Spring")
)
```

Checking for missing expansions

This section will look at where estimated (expanded) tags are missing from the SAR data queried. Users can specify how they would like recovery data grouped to look for missing expansions of CWT data. Year is included by default. Missing expansions are important because often SAR analysis relies on expanded recoveries and years and locations that report raw recoveries only many bias SAR analysis. This function will write out a csv file to look at what % of raw recoveries are unexpanded by release group (based on the release group variables the user specifies). If a lot of the raw recoveries are unexpanded, consider further filtering your dataset to exclude years/recoveries types/release locations plagued by this problem.

```
missing_expansions(
  datafile = "CowlitzSpChk_CWT_data.xlsx",
  MPRfilters = c("High Seas","Juvenile Sampling","Terminal Sport"),
  CWT_status = "CWT",
  mark = "Ad",
  min_ocean_age = 2,
  min_total_age = 3,
  max_ocean_age = 5,
  max_total_age = 6,
  rel_groups = c("hatchery","brood_year"),
  rec_groups = c("MPR_Groupings")
)
```

Prepare data for SAR analysis (v1: SAR analysis by individual tag code)

Here, we will take the further refine our filtering, and process our initial raw data pull from RMIS into an easily digestible format for analysis. This function can be used to assemble data by groups of tag codes, or at the individual tag code level. Below the function is set up to conduct a grouped analysis. This will write the final ready-for-analysis csv file to your "results" directory.

```
dat<-preparedata(
  datafile ="CowlitzSpChk_CWT_data.xlsx",
  MPRfilters = c("High Seas","Juvenile Sampling","Terminal Sport"),
  CWT_status = "CWT",
  mark = "Ad",
  min_ocean_age = 2,
  min_total_age = 3,
  max_ocean_age = 5,
  max_total_age = 6,
  adjust_expanded_tags_for_terminal_harvest = T,
  terminal_harvest_file="CowlitzSpCkHarvestRate_byage_10.20.2020.csv",
  set_expanded_to_raw_when_missing=T, #must be T if expanding escapement for terminal harvest!!!
  create_age_season_group = F, #must be T if expanding escapement for terminal harvest!!!
  release_group_DOYs = NULL,
  release_group_cats= NULL,
  rel_groups<-c("hatchery","stock","release_site","brood_year","tag_code","first_release_month","first_
  rec_groups<-c(NULL),
  writefile = "Yes",
  filename = "CowlitzSpChk_SAR_Analysis_bytag_"
)
```

Run analysis (v1: SAR analysis by individual tag codes)

This code will run the SAR analysis using JAGS, where SAR is estimated via a gam. The model structure is currently defined in the "runanalysis" function and may need to be edited to accomodate your analysis of interest.

```
using("MASS","R2jags","RColorBrewer","mgcv")
run_analysis_cowlitz(datafile = "CowlitzSpChk_SAR_Analysis_bytag_Data_for_Analysis.csv",
             bygrouportagcode = "tagcode",
             grouplist=NA,#list groups for group analysis
             options=NA #this is if you wanted to compare total returns as a function of different mixt
)
```

Prepare data for SAR analysis (v2: SAR analysis by grouped tag codes)

Here, we will take the further refine our filtering, and process our initial raw data pull from RMIS into an easily digestible format for analysis. This function can be used to assemble data by groups of tag codes, or at the individual tag code level. Below the function is set up to conduct a grouped analysis. A version by tag code will follow.This will write the final ready-for-analysis csv file to your "results" directory.

```
dat<-preparedata(
  datafile ="CowlitzSpChk_CWT_data.xlsx",
  MPRfilters = c("Terminal Sport","High Seas","Juvenile Sampling"),
  CWT_status = "CWT",
  mark = "Ad",
  min_ocean_age = 2,
  min_total_age = 3,
  max_ocean_age = 5,
  max_total_age = 6,
  adjust_expanded_tags_for_terminal_harvest = F,
  terminal_harvest_file="CowlitzSpCkHarvestRate_byage_10.20.2020.csv",
  set_expanded_to_raw_when_missing=T, #must be T if expanding escapement for terminal harvest!!!
  create_age_season_group = F, #must be T if expanding escapement for terminal harvest!!!
  release_group_DOYs = NULL,
  release_group_cats= NULL,
  rel_groups<-c("hatchery","stock","release_site","brood_year","tag_code","first_release_month","first_
  rec_groups<-c(NULL),
  writefile = "Yes",
  filename = "CowlitzSpChk_SAR_Analysis_bytag_"
)
```

Run analysis (v2: SAR analysis by grouped tag codes)

This code will run the SAR analysis using JAGS, where logit(SAR) follows a multivariate normal random walk [INSERT MORE DETAILS OF GROUP ANALYSIS HERE]

```
# run_analysis_cowlitz(dat = "CowlitzSpChk_SAR_Analysis_bygroup_Data_for_Analysis.csv",
#              bygrouportagcode = "group",
#              grouplist=NA,#list groups for group analysis
#              options=NA #this is if you wanted to compare total returns as a function of different mi
# )
```

Evaluating Residualism

We will now conduct a separate analysis to look at residualism of released hatchery smolts, comparing smolt age by tag group reported to RMIS with the observed freshwater age from CWT recoveries in the hatchery, sport fishery, and on the spawning grounds

```
dat<-data.frame(read_csv(file.path(getwd(),"results","CowlitzSpChk_SAR_Analysis_bytag_Data_for_Analysis
CWT_length_age<-data.frame(read_xlsx(path=file.path(getwd(),"supplementaldata","Cowlitz Chinook scale a
results<-data.frame(dat%>%left_join(CWT_length_age,by=c("tag_code"))%>%
  filter(!is.na(AgeRead) & !is.null(AgeRead) & AgeRead!=9)%>%
  mutate(FW_age_scales=substr(AgeRead,2,2),season=ifelse(first_release_month<6,"spring",ifelse(first_rel
  group_by(release_age,season,FW_age_scales)%>%
  summarize(count=n())%>%
  mutate(freq = count / sum(count)))
write.csv(results,paste(getwd(),"/results/Residualism_Scales_vs_CWT_FWage.csv",sep=""))
kable(results,caption="Table 1. Correspondence between freshwater scale age and CWT age for Cowlitz hatc
```

Table 1: Table 1. Correspondence between freshwater scale age and CWT age for Cowlitz hatchery spring Chinook.

release_age	season	FW_age_scales	count	freq
1	fall	1	1	0.0172414
1	fall	2	57	0.9827586
1	spring	2	1	1.0000000
1	summer	1	1	0.2000000
1	summer	2	4	0.8000000
2	spring	1	52	0.0197044
2	spring	2	2587	0.9802956