Simulated dataset for WKRDB_EST truc

Laurent Dubroca

02/10/2019

Introduction

- simulation of a fish population fished by a fleet of vessels. Three gears with different selectivity curves. Then data are generated according to the sampling scheme used in the french at-sea sampling scheme.
- based on the simulated data, a dirty (and probably buggy) hierarchy one set of files is generated.

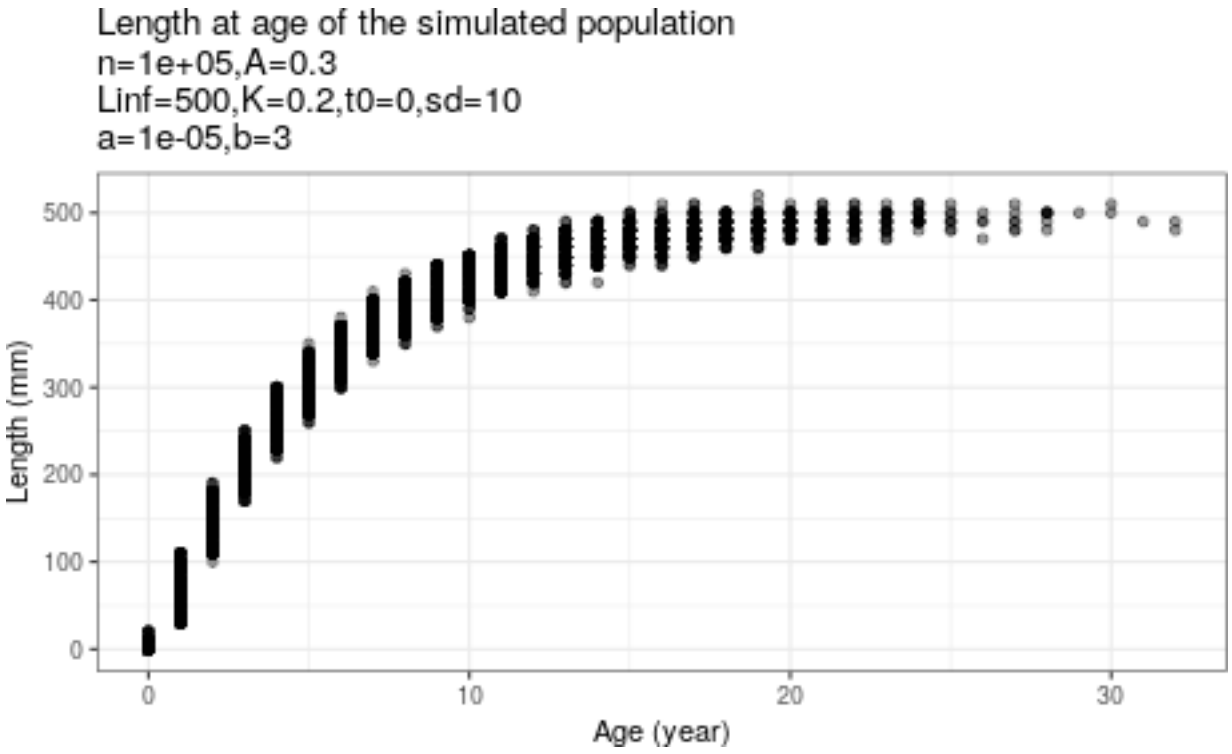
```
#knitr option
operationnel<-TRUE
knitr::opts_chunk$set(echo=FALSE,
              warning=!operationnel,
              message=!operationnel,
              fig.height=4,
              progress=!operationnel,
              verbose=!operationnel,
              include=TRUE,dev='png',autodep=FALSE)
#package
library(dplyr)
library(ggplot2)
library(sampling)
library(FSA)
library(FSAsim)
library(purrr)
##library(openxlsx);library(mailR)
library(pander);library(captioner)
#local fct
#source("credo_fct.R")
#initialise les fonctions pour légender tables et figures
tabcap <- captioner(prefix = "Table")
figcap <- captioner(prefix = "Figure")

#general parameters
nbpopfish<-100000 #number of fish in the virtual population
natmortpopfish<-0.3 # the natural mortality of the virtual population
vbLinf <- 500  #von berta: Linf value of the virtual fish
vbK <- 0.2#von berta: K value of the virtual fish
vbt0 <- 0#von berta: t0 of the virtual fish
vbsigma <- 10#von berta sd around age for length distribution
rtp<-data.frame(sciname="Virtus piscem",a=0.01/1000,b=3)

#fix randomness with the number of the Beast
set.seed(666)
```

A virtual fish population

Simulation of a virtual fish population.

```r
#a population of fish A
popage<-simAges(N0=nbpopfish,A=natmortpopfish,incl.zero=T)
#their lengths
poplen <- simLenFromAge(popage,Linf=vbLinf,K=vbK,t0=vbK,sigma=vbsigma)
#the population dataframe (filtering out negative length if any)
popfish <- data.frame(age=popage,len=poplen)%>%filter(len>0)
#adding weight
popfish$w<-rtp$a*(popfish$len/10)^rtp$b
#trunc the fish size
popfish$len <- lencat(popfish$len,w=10)
#some text summarizing the simulation parameters
txtparam1<-paste0("n=",nbpopfish,",A=",natmortpopfish)
txtparam2<-paste0("Linf=",vbLinf,",K=",vbK,",t0=",vbt0,",sd=",vbsigma)
txtparam3<-paste0("a=",rtp$a,",b=",rtp$b)#,",t0=",vbt0,",sd=",vbsigma)
```

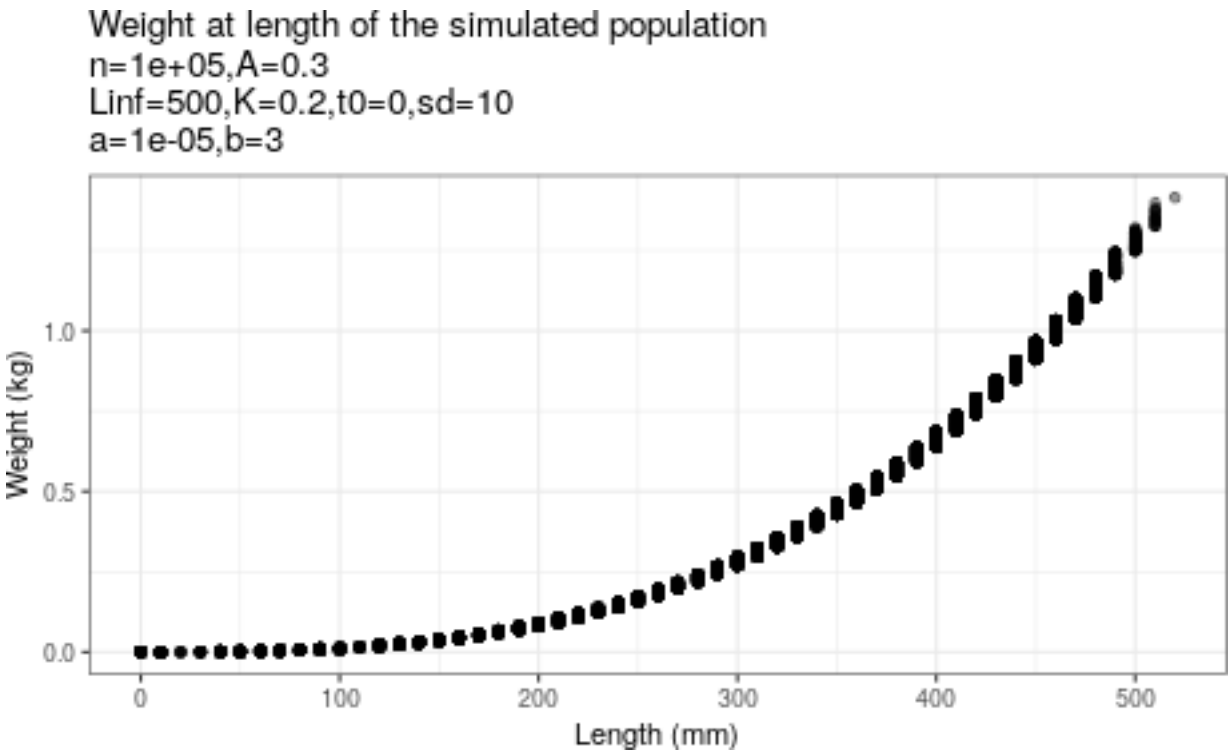
```r
p1<-ggplot(popfish,aes(x=age,y=len))+geom_point(alpha=.4)+
    theme_bw()+xlab("Age (year)")+ylab("Length (mm)")+
    ggtitle(paste0("Length at age of the simulated population\n",
                txtparam1,"\n",txtparam2,"\n",txtparam3))
print(p1)
```




```r
cat(figcap(name="popsimplt1",caption="Length at age in the virtual population"))
```

Figure 1: Length at age in the virtual population

```r
p1<-ggplot(popfish,aes(x=len,y=w))+geom_point(alpha=.4)+
    theme_bw()+xlab("Length (mm)")+ylab("Weight (kg)")+
    ggtitle(paste0("Weight at length of the simulated population\n",
                txtparam1,"\n",txtparam2,"\n",txtparam3))
print(p1)
```




```
cat(figcap(name="popsimplt1",caption="Weight at length in the virtual population"))
```

Figure 1: Length at age in the virtual population

```
p2<-ggplot(tidyr::gather(popfish),aes(value))+geom_histogram()+#binwidth=c(1,10))+
    facet_wrap(~key,scale="free")+
    theme_bw()+ylab("Count")+xlab("Value")+
    ggtitle(paste0("Age and length distribution of the simulated population\n",txtparam1,"\n",txtparam2)
print(p2)
```

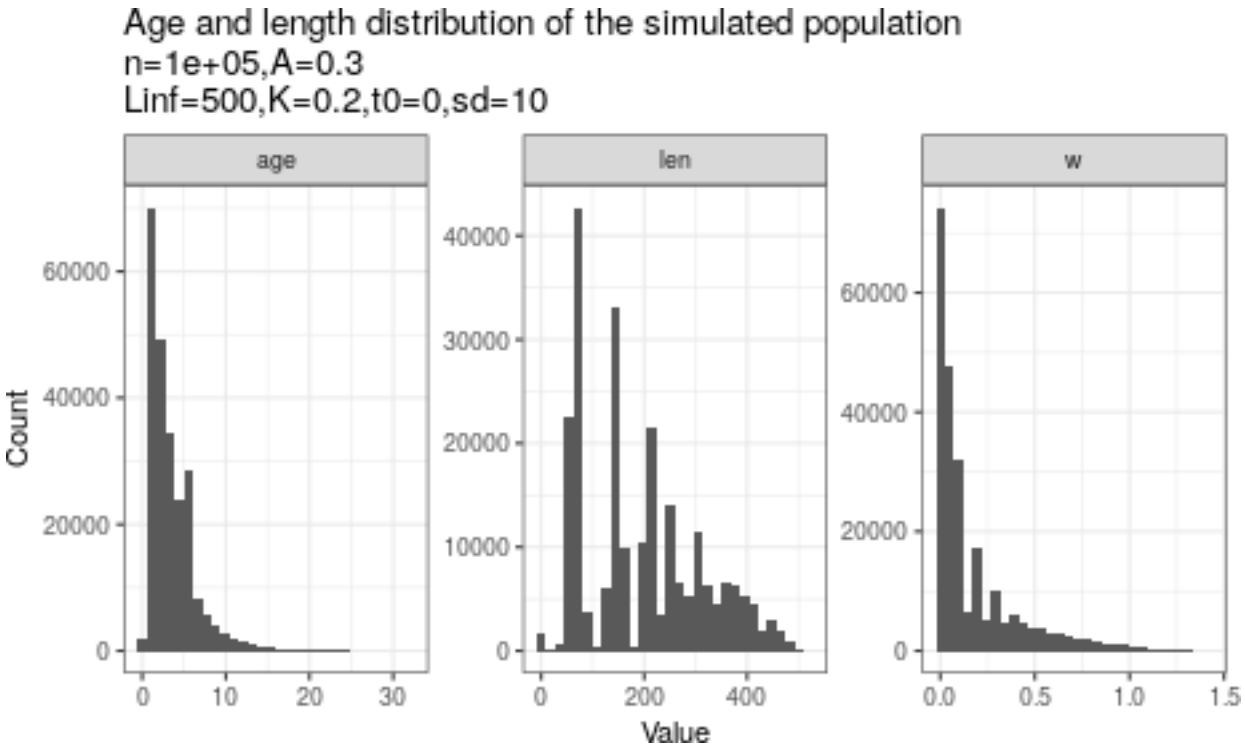
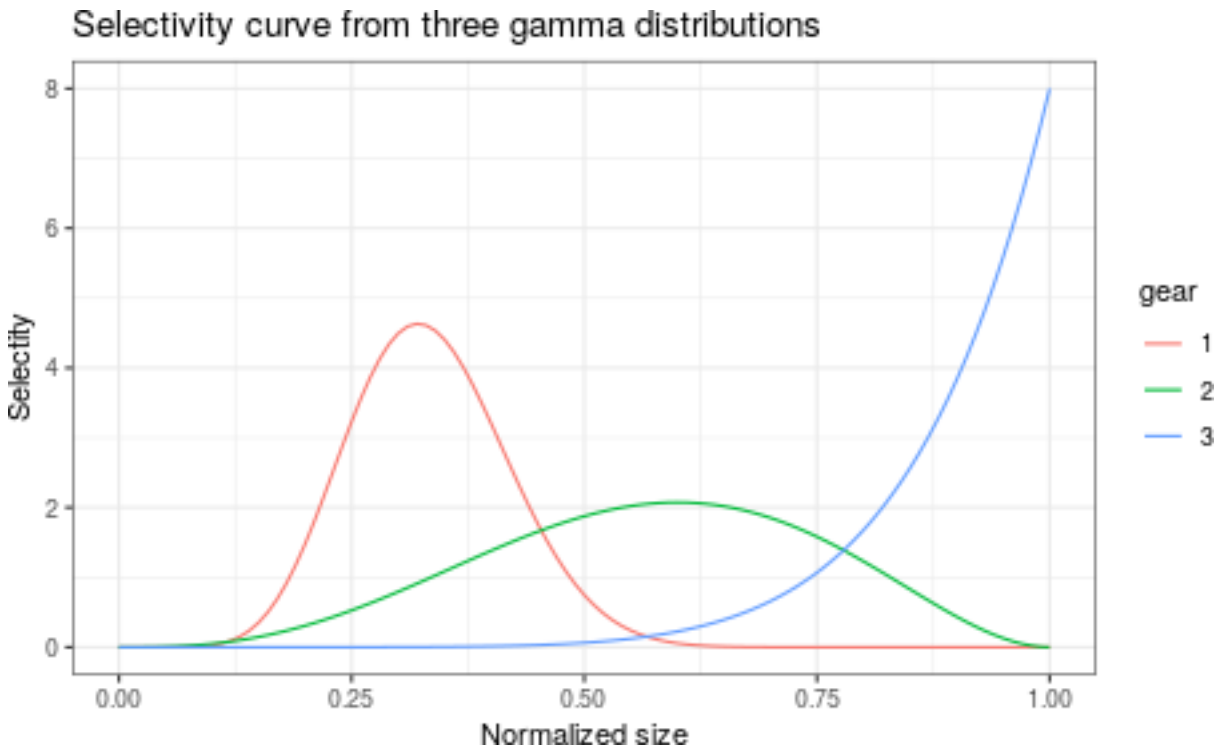


Figure 2: Length distribution of the virtual population

A virtual fleet targeting a virtual fish population

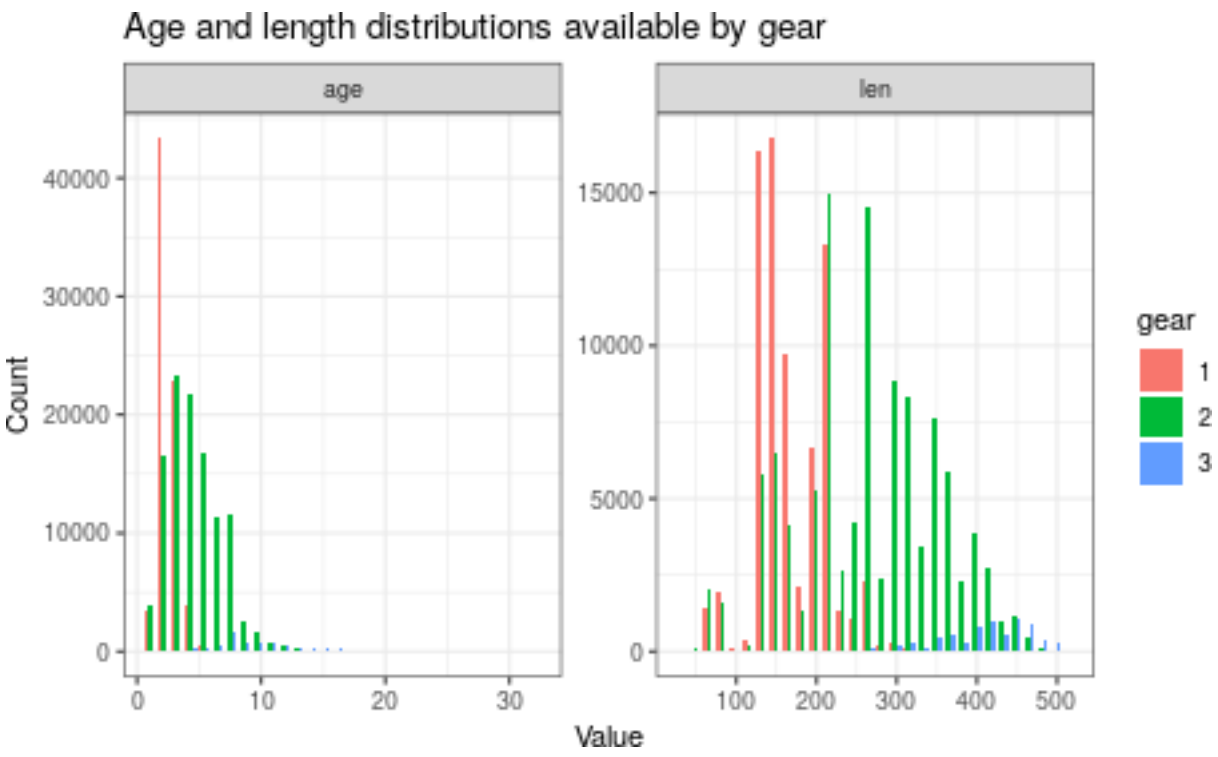
A gear is defined by its selectiviy curve. This curve is supplied using a beta distribution. A graphical representation of the selectivity curves is provided.

```
selex1<-data.frame(x=seq(0,1,0.01))%>%mutate(sel=dbeta(x,shape1=10,shape2=20),gear="1")
selex2<-data.frame(x=seq(0,1,0.01))%>%mutate(sel=dbeta(x,shape1=4,shape2=3),gear="2")
selex3<-data.frame(x=seq(0,1,0.01))%>%mutate(sel=dbeta(x,shape1=8,shape2=1),gear="3")
selex<-rbind(selex1,selex2,selex3)
p1<-ggplot(selex,aes(x=x,y=sel,color=gear))+
    geom_path()+
    theme_bw()+ylab("Selectity")+xlab("Normalized size")+
    ggtitle(paste0("Selectivity curve from three gamma distributions"))#\n",txtparam1,"\n",txtparam2))
print(p1)
```



Population by gears + trips:

```
#population accessible to each trip
lengear1<-simLenSelectP(popfish$len, alpha = 10, beta = 20, max.height = 1, show = FALSE)
lengear2<-simLenSelectP(popfish$len, alpha = 4, beta = 3, max.height = 1, show = FALSE)
lengear3<-simLenSelectP(popfish$len, alpha = 8, beta = 1, max.height = 1, show = FALSE)
popfish1<-popfish%>%filter(lengear1)%>%mutate(gear="1")
popfish2<-popfish%>%filter(lengear2)%>%mutate(gear="2")
popfish3<-popfish%>%filter(lengear3)%>%mutate(gear="3")
popfishgear<-rbind(popfish1,popfish2,popfish3)
p1<-ggplot(tidyr::gather(popfishgear,key="key",value="value",age:len),
       aes(value,fill=gear,group=gear))+
    geom_histogram(position="dodge")+
    facet_wrap(~key,scale="free")+
    theme_bw()+ylab("Count")+xlab("Value")+
    ggtitle(paste0("Age and length distributions available by gear"))
print(p1)
```




```
#generate some trips
#a fct to generate 1 haul, then 1 trip, then 1 vessel
do1haul<-function(df){
    sample_n(df,sample(50:450,1))%>%
        mutate(idhaul=sub("\\.","",format(Sys.time(),"%s%OS3")))
}
do1trip<-function(df,nbhaul=sample(1:10,1)){
    do.call("rbind",rerun(nbhaul,do1haul(df)))%>%
        mutate(idtrip=sub("\\.","",format(Sys.time(),"%s%OS3")))
}
do1vessel<-function(df,nbtrip=sample(1:50,1)){
    do.call("rbind",rerun(nbtrip,do1trip(df)))%>%
        mutate(idvessel=sub("\\.","",format(Sys.time(),"%s%OS3")))
}


#generate 100 trip for popfish1
vess1<-do.call("rbind",rerun(50,do1vessel(popfish1))) %>%mutate(gear="1")
vess2<-do.call("rbind",rerun(30,do1vessel(popfish1))) %>%mutate(gear="2")
vess3<-do.call("rbind",rerun(10,do1vessel(popfish1))) %>%mutate(gear="3")
popvess<-rbind(vess1,vess2,vess3)%>%ungroup()%>%mutate(pk=row_number())
#select vessel and then trip and then systematic on haul 1 on 3
vesselection<-popvess%>%select(gear,idvessel)%>%distinct()%>%
    group_by(gear)%>%
    mutate(n=n_distinct(idvessel))%>%
    sample_frac(.1,weight=n)#unique(popvess$vessid)
#select n trip by vessel
tripselection<-popvess%>%select(idvessel,idtrip)%>%distinct()%>%
    filter(idvessel%in%vesselection$idvessel)%>%
    group_by(idvessel)%>%
```

```
    mutate(n=n_distinct(idtrip))%>%
    sample_n(1)
#select 1 haul every 3 haul
haulselection<-popvess%>%select(idvessel,idtrip,idhaul)%>%distinct()%>%
    filter(idtrip%in%tripselection$idtrip)%>%
    group_by(idtrip)%>%
    mutate(id=row_number())%>%
    filter(id%in%seq(1,10,2))%>%select(-id)
#select n length by haul and add a samp id
selvess<-left_join(haulselection%>%ungroup(),popvess)%>%
    group_by(idhaul)%>%
    sample_n(30)%>%mutate(idsamp=group_indices())
#pop file with sampling id
popvess<-full_join(popvess,selvess%>%select(pk,idsamp)%>%distinct())
saveRDS(popvess,file="./data/popvess.rds")
```

Simulation to RDBES data

From the simulated data, a hierarchy 1 is generated (some tables are omitted).

```
#####################################################
#reverse methods
#start from lower hierarchy then build up information needed in each table
FM<-popvess%>%filter(!is.na(idsamp))%>%group_by(idsamp,FMclass=len)%>%
    summarise(FMnumberAtUnit=n())%>%
    ungroup()%>%
    transmute(FMid=row_number(),
          SAid=idsamp,
          FMrecType="FM",
          FMclass,
          FMnumberAtUnit,
          FMtype="length",
          FMMeasurementEquipement="",
          FMaccuracy="",
          Fmsampler="")

#####################################################
#SA
tmp<-popvess%>%group_by(idhaul)%>%mutate(haulw=sum(w))%>%
    ungroup()%>%group_by(idsamp,haulw)%>%summarise(sampw=sum(w))%>%ungroup()
SA<-tmp%>%transmute(SAid=idsamp,
          SAparentID=NA,
          SSid="SSsppA",
          SArecType="SA",
          SAnatCode="",
          SAstratification="N",
          SAstratum="U",
          SAsppCode="Genus species",
          SAcommSpp="AAA",
          SApres="",
          SAcatchCat="Lan",
          SAlandCat="HUC",
```

```
          SAcommCatScl="",
          SAcommCat="",
          SAsex="U",
          SAunitType="Haul",
          SAtotalWtLive=haulw,
          SAsampWtLive=sampw,
          SAtotal=1,
          SAsampled=1,
          SAprob=NA,
          SAselectMeth="Census",
          SAlowHierarchy="FM",
          SAsampler="Observer",
          SAnoSampReasonFM="",
          SAnoSampReasonBV="",
          SAtotalWtMes=haulw,
          SAsampWtMes=sampw,
          SAconFacMesLive=1,
          SAspecState="")

#species selection
tmp<-popvess%>%filter(!is.na(idsamp))%>%select(idhaul)%>%distinct()
SS<-tmp%>%transmute(SSid="SSsppA",
           LEid=NA,
           FOid=idhaul,
           SLid="SLsppA",
           SSrecType="SS",
           SSstratification="U",
           SSobsActTyp="Haul",
           SScatchCat="Lan",
           SSobsTyp="visual",
           SSstratum="U",
           SSclustering="No",
           SSclusterName="U",
           SSsampler="",
           SSsppListName="",
           SStotal=1,
           SSsampled=1,
           SSselectMeth="Census",
           SSselectMethCluster="U",
           SStotalClusters="",
           SSsampledClusters="",
           SSprobCluster="",
           SSnoSampReason="")
############################################################
#FO
tmp<-popvess%>%group_by(idtrip)%>%mutate(check=sum(idsamp,na.rm=T))%>%
    ungroup()%>%filter(check>0)%>%
    select(idhaul,idtrip,idsamp)%>%distinct()%>%
    group_by(idtrip)%>%mutate(tothaul=n_distinct(idhaul),samphaul=n_distinct(idsamp)-1) %>%
    ungroup()%>%
    select(idhaul,idtrip,tothaul,samphaul)%>%distinct()%>%
    group_by(idtrip)%>%mutate(n=row_number())%>%ungroup()
FO<-tmp%>%transmute(FOid=idhaul,
```

```
            FTid=idtrip,
            SDid="",
            FOrecType="FO",
            FOstratification="N",
            FOhaulNum=n,
            FOstratum="U",
            FOclustering="No",
            FOclusterName="",
            FOsampler="",
            FOaggLev="H",
            FOval="V",
            FOcatReg="Lan",
            FOstartDate="",
            FOstartTime="",
            FOendDate="",
            FOendTime="",
            FOdur="",
            FOstartLat="",
            FOstartLon="",
            FOstopLat="",
            FOstopLon="",
            FOecoZone="",
            FOarea="",
            FOstatRect="",
            FOsubRect="",
            FOdep="",
            FOwaterDep="",
            FOnatCat="",
            FOmetier5="",
            FOmetier6="",
            FOgear="",
            FOmeshSize="",
            FOselDev="",
            FOselDevMeshSize="",
            FOtarget="",
            FOobsCo="Ha",
            FOtotal=tothaul,
            FOsampled=samphaul,
            FOprob="",
            FOselectMeth="SYSS",
            FOselectMethCluster="",
            FOtotalClusters="",
            FOsampledClusters="",
            FOprobCluster="",
            FOnoSampReason="")

##############################################################
#FT
tmp<- popvess%>%group_by(idvessel)%>%mutate(nbtrip=n_distinct(idtrip))%>%
    group_by(idvessel,idtrip)%>%mutate(nbhaul=n_distinct(idhaul))%>%
    group_by(idvessel,idtrip,idhaul)%>%mutate(nbsamp=n_distinct(idsamp)-1)%>%
    ungroup()%>%
    select(idvessel,idtrip,nbtrip,nbhaul,nbsamp)%>%distinct() %>%
```

```
    filter(nbsamp>0)


FT<-tmp%>%transmute(FTid=idtrip,
            OSid="",
            VSid=idvessel,
            VDid=idvessel,
            SDid="",
            FOid="",
            TEid="",
            FTrecType="FT",
            FTnatCode="",
            FTstratification="N",
            FTstratum="U",
            FTclustering="No",
            FTclusterName="U",
            FTsampler="",
            FTfoNum=nbhaul,
            FTdepLoc="",
            FTdepDate="",
            FTdepTime="",
            FTarvLoc="",
            FTarvDate="",
            FTarvTime="",
            FTtotal=nbtrip,
            FTsampled=nbsamp,
            FTprob="",
            FTselectMeth="random",
            FTselectMethCluster="",
            FTtotalClusters="",
            FTsampledClusters="",
            FTprobCluster="",
            FTnoSampReason=""
            )

#VS
tmp<- popvess%>%group_by(gear)%>%mutate(nbvessel=n_distinct(idvessel))%>%
    group_by(gear,idvessel)%>%mutate(nbsamp=n_distinct(idsamp)-1)%>%
    ungroup()%>%
    filter(nbsamp>0)%>%
    select(gear,idvessel,nbvessel,nbsamp)%>%distinct()

VS<-tmp%>%transmute(VSid=idvessel,
        SDid=2,
        VDid=idvessel,
        TEid="",
        VSrecType="VS",
        VSstratification="Y",
        VSstratum=paste("Gear",gear),
        VSclustering="N",
        VSclusterName="U",
        VSsampler="Someone",
        VStotal=nbvessel,
```

```
        VSsampled=nbsamp,
        VSprob=.1,
        VSselectMeth="random",
        VSselectMethCluster="",
        VStotalClusters="",
        VSsampledClusters="",
        VSprobCluster="",
        VSnoSampReason="")
DE<-data.frame(DEid=1,
        DErecType="DE",
    DEsampScheme="sim H1",
    DEyear=2054,
    DEstratum=c("Stratified by fleet"),
    DEhierarchyCor=c("Yes"),
    DEhierarchy="H1")
SD<-data.frame(SDid=1,
           DEid=1,
           SDrecType="SD",
           SDctry="XXX",
           SDinst="Fishery Agency")


simH1<-list(DE,SD,VS,FT,FO,SS,SA,FM)
saveRDS(simH1,file="simH1.rds")
```