A COMPARISON OF AGE SLICING AND STATISTICAL AGE ESTIMATION FOR MEDITERRANEAN SWORDFISH (*Xiphias gladius*).

Laurence T. Kell[1] & Alexander Kell[2]

SUMMARY

Catch-at-size (CAS) data are important inputs for many stock assessment methods. However, for methods such as Virtual Population Analysis (VPA) catch-at-size are often first converted into catch-at-age (CAA) by age slicing using a deterministic growth model. In this paper an alternative statistically based approach, which assumes that the CAS is composed of a mixture of length frequency distributions, is used to estimate age composition. In this example based on Mediterranean swordfish it is shown that age slicing underestimated both the proportion of younger fish in the catch and uncertainty in the catch-at-age estimates.

KEYWORDS

Age slicing ,catch curve analysis, FLR, Mediterannean, Monte Carlo simulation, reference points, statistical age estimation, swordfish.

1. Introduction

Catch-at-size (CAS) data are important inputs for many stock assessment methods. However, for use by methods such as Virtual Population Analysis (VPA) catch-at-size often have first to be converted into catch-at-age (CAA). This is usually performed, prior to running an assessment, by age slicing where a deterministic growth model is used to divide the length frequency data into length bins. An alternative statistically based approach is to assume that the CAS is composed of a mixture of length frequency distributions and to estimate the parameters by the method of maximum likelihood.

In this paper the two approaches, *age slicing* and *statistical age estimation,* are compared using catch-at-size data for Mediterranean swordfish. A catch curve analysis is performed to compare total mortality estimated from the two sets of catch-at-age data and then reference points based on based on selections patterns derived from the two sets of catch-at-age data are compared. Following this analysis uncertainty in the estimates of catch-at-age obtained from the two methods is evaluated.

2. Material and Methods

The data used were Mediterranean swordfish catch-at-size from the ICCAT database up to and including 2008 **figure 1**; growth was assumed to follow a Von Bertalanffy curve (Tserpes and Tsimenides, 1995). These data are available from www.iccat.int/???

All modelling was performed in R using FLR (www.flr-project.org, Kell et al. 2007) and statistical fitting of length distributions at age was performed using the mixdist package (Manning ?)

Age slicing predicts age from length using the inverse of the Von Bertalanffy growth curve i.e.

$$a = t_0 - \log(1-l)/K$$

Where if $l(a) \leq l < l(a+1)$ then the assumed age is α.

---

[1] ICCAT Secretariat, C/Corazón de María, 8. 28002 Madrid, Spain; Laurie.Kell@iccat.int; Phone: 34 914 165 600 Fax: 34 914 152 612.

[2] El Berzalejo

Statistical fitting in contrast assumes that a length-frequency distribution is a weighted sum of k component densities i.e.

$$g(x|\boldsymbol{\pi}, \mu, \boldsymbol{\sigma}) = \boldsymbol{\pi_1} f(x|\pi_1, \mu_1, \sigma_1) + \cdots + \boldsymbol{\pi_1} f(x|\pi_k, \mu_k, \sigma_k)$$

where the parameters are the mixing proportions (π) the means (μ) and standard deviations ($\boldsymbol{\sigma}$) of the component distributions. Within the mixdist package components can be either normal, lognormal, gamma, exponential, Weibull, binomial, negative binomial or Poisson distributions.

Following generation of the catch-at-age matrices catch curve analyses were performed to estimate total mortality (Z) and a yield per recruit analysis conducted to compare reference points derived from *age slicing* and *statistical age estimation*.

Estimates of uncertainty in numbers-at-age were estimated for *age slicing* using Monte Carlo simulation and compared with uncertainty derived from the statistical fits.

Catch curves, where log(catch) is plotted against age, is a simple method to explore patterns in (Z). As with VPA it is assumed that the population is closed to emigration and immigration and that

$$N_{a,t} = N_{a-1,t-1} \, e^{-Z_{a-1,t-1}}$$

Where $N_{a,t}$ and $Z_{a,t}$ are numbers and total mortality-at-age, and $N_{0,t}$ is the initial population size of a cohort.

Catch-at-age is then a function of numbers-at-age assuming that a constant proportion of the population is vulnerable to the fishery (q_a)

$$C_{a,t} = q_a N_{a,t}$$

Taking logs of each side gives the linear relationship

$$log(C_t) = log(qN_0) - Z_t$$

The slope (Z_t) provides an estimate of total mortality within a year class if catchability (q_a) is constant over ages. However, this is generally violated as catchability may initially increase and then decline at older ages. In the case of a flat topped selection pattern the catch curve will show an ascending left limb, a domed middle portion, and a descending right limb. The ascending left limb represents age-classes of fish that are not fully recruited to the fishery and the width of the domed middle section provides an indication of the rate of recruitment to the fishery. Once the right hand limb is identified its slope can provide an estimate of Z. Doing this by cohort allows patterns over time to be evaluated. However, changes in selectivity-at-age and effort between years may well mean that important changes may be obscured.

A yield per recruit analysis was performed to evaluate the effect on estimates of F_{Max} and $F_{0.1}$ of the two methods of age determination. This was done by adjusting selectivity-at-age obtained from the 2010 assessment (ref ?) based on age slicing so that catch proportions corresponded to those of statistically estimated catch-at-age matrix and then recalculating the yield per recruit curve and reference points.

The statistical age estimation method provides standard errors for all estimated parameters however age slicing does not. Therefore a Monte Carlo simulation was performed, where the observed catch-at-size was re-sampled and catch proportions re-estimated, allowing coefficients of variation for proportions at age to be estimated.

3. Results

Catch-age size data are presented in **figure 2**, observed length frequency distributions are plotted in blue, the expected length at each birthday are indicated by the vertical lines, the modes of the fitted length distributions-at-age are plotted in red and the fitted mixture distribution in green. In general the fitted distribution shows good agreement with the observations, particularly for age 0 through to 3; subsequently the modes become more

difficult to distinguish. The relative proportions at age are compared in **figure 3**, the statistical age slicing results in a higher proportion of catches occurring at younger ages (i.e. 0,1 and 2).

In order to see the effect on mortality a catch curve analysis was conducted, **figure 4,** where the slope provides an estimate of Z. The log catch ratios are plotted against age and year in **figures 5 & 6** respectively.

The effect of the change in selection pattern is evaluated by inspection of the yield per recruit curves (**figure 7**). The blue curve is from the 2010 assessment and in the red curve selectivity-at-age was adjusted so that the mean catch proportion-at-age corresponded to the statistically estimated catch-at-age matrix.

Coefficients of variation for catch proportion-at-age from the statistical age estimate, derived from the fit and the Monte Carlo simulation to obtain CVs for age slicing for 2009 are presented in **table 1**.

4. Discussion

Two important differences were seen in the catch-at-age derived from slicing and statistical estimation i.e. numbers-at-age of the younger ages, which were more numerous, were underestimated by slicing and uncertainty in proportions of the older ages, where there was considerable overlap in length-at-age, was much greater in the statistical age estimates.

The results raise important questions for the use of the catch-at-age generated by age slicing for use by virtual population analysis (VPA). Since estiamates of stock status, fishing mortality and reference points will be biased. Also importantly VPA assumes that catch-at-age is known without error, therefore advice based upon probabilities estimates, such as Kobe plots, are likely to be over optimistic.
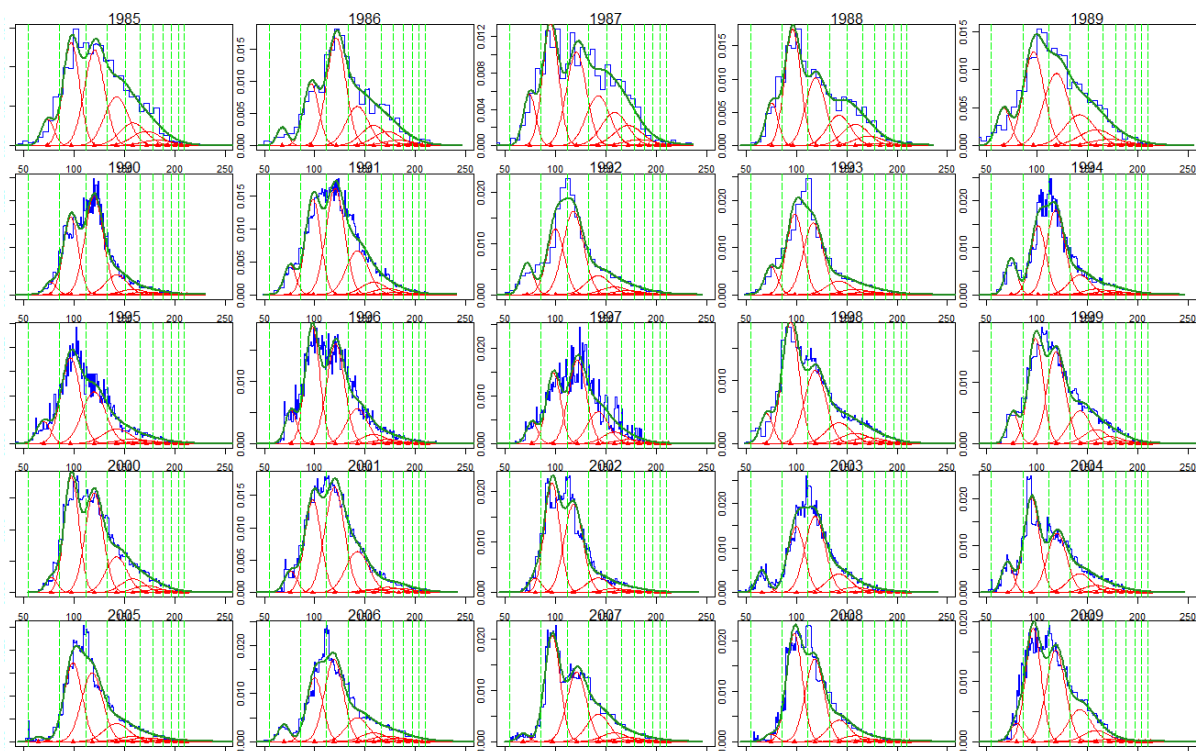

5. References

Kell, L.T., Mosqueira, I., Grosjean, P., et al. (2007) FLR: an open-source framework for the evaluation and development of management strategies. ICES Journal of Marine Science 64:640-646

Manning (http://www.math.mcmaster.ca/peter/mix/mixdist.pdf).

Tserpes and Tsimenides, 1995
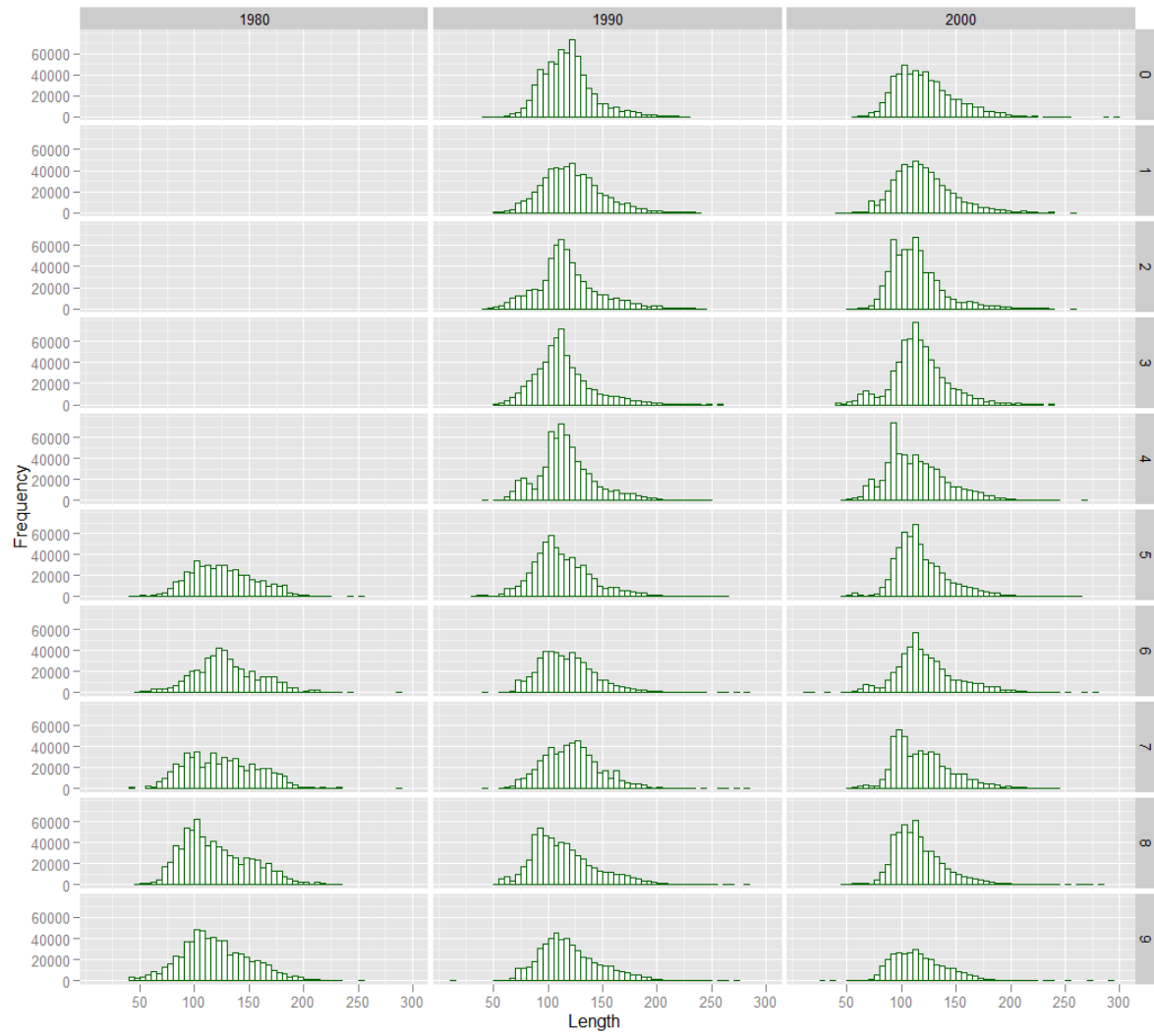
**Figure 1.** Observer length frequency distributions by decade.

Figure 2. Simulated stochastic length slicing; blue denotes observed size frequency, red the fitted length distributions to each age, green solid line the fitted size distribution and green vertical hatched lines length at each birthday.
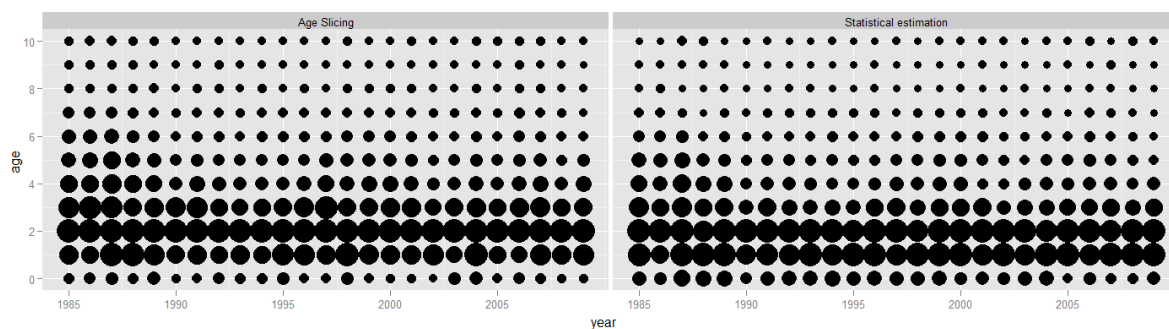

Figure 3. Catch-at-age matrices scaled by maximum catch number-at-age within a year, blue line is a loess smoother fitted to all years.
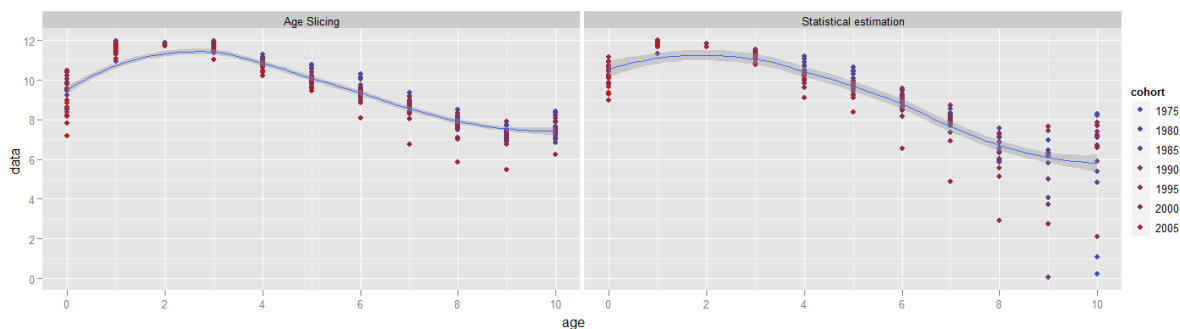
Figure 4. Catch curves, i.e. logged catch-at-age plotted against year, blue line is a loess smoother fitted to all years.
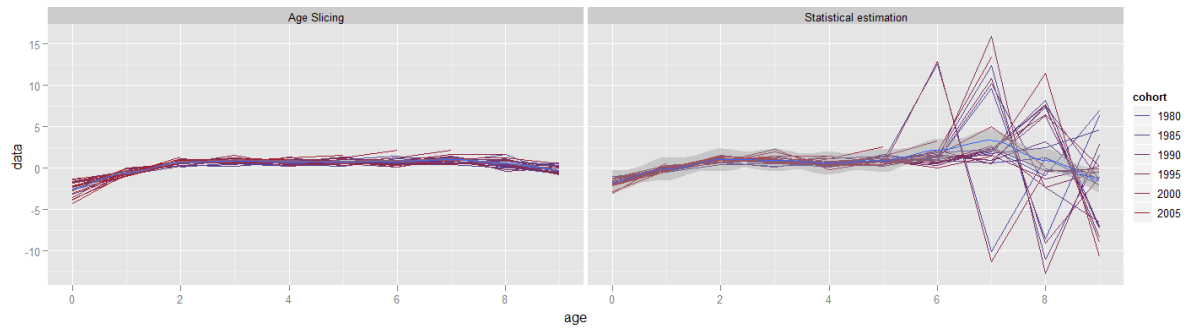
Figure 5. Log catch ratios plotted against age, curves represent cohorts, blue line is a loess smoother fitted to all years.
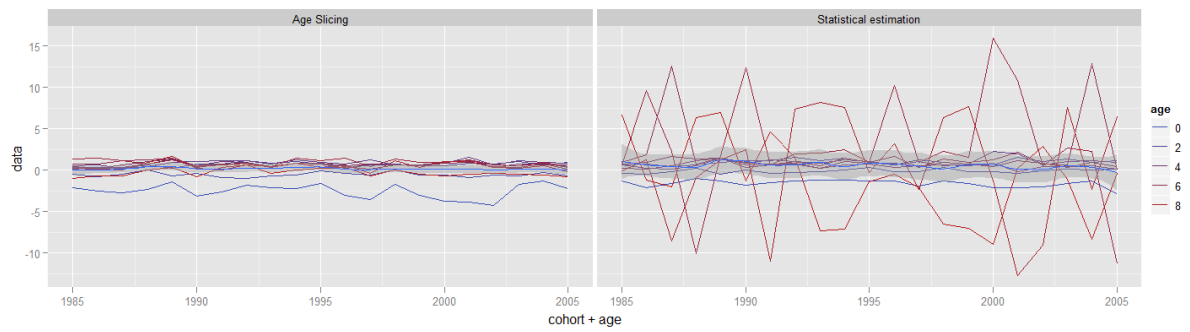
Figure 6. Log catch ratios plotted against year, curves represent cohorts, blue line is a loess smoother fitted to all years.
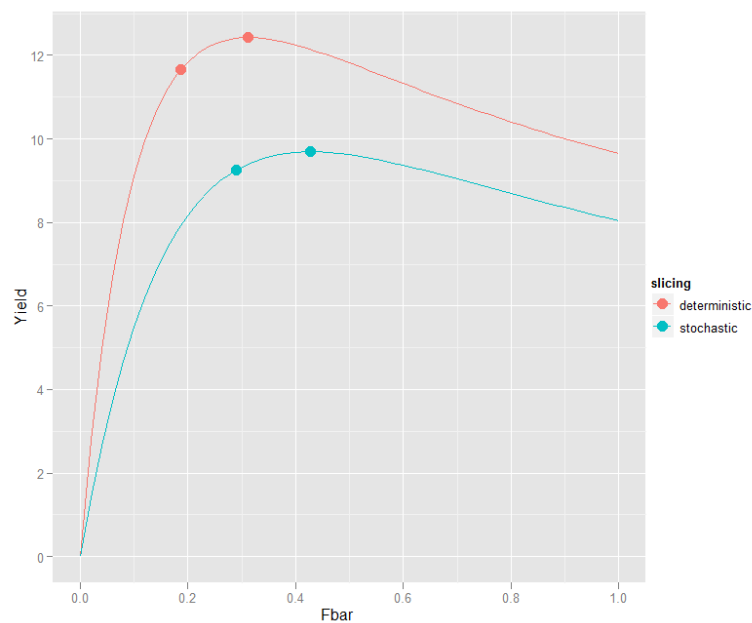
Figure 7. Yield per recruit curves, with points corresponding to $F_{0.1}$ & F_{Max}

Table 1. Coefficient of variation in proportion of catch-at-age from statistical age estimation and deterministic age slicing for 2009.

Age	Statistical Age Estimation	Age Slicing
0		22%
1	1%	12%
2	0%	12%
3	2%	15%
4	6%	18%
5	49%	19%
6	463%	27%
7	2795%	29%
8	180284000%	32%
9	7400%	36%
10	27%	15%

6. Code

This section contains all the code to perform the analyses in the paper. There are 9 **"Code Chunks"** that load the necessary R package, data and functions, perform the age slicing and statistical ageing, then calculate catch proportions and undertake catch and Catch Ratio analysis. Following which reference points are calculated and a Monte Carlo simulation is perfprmed to estimate coefficients of variation for catch proportions estimated from age slicing.

i)   Initial

Three R package or libraries are required for the analysis; FLCore and FLBRP for fisheries methods, mixdist for statistical estimation of proportions-at-age and ggplot2 for plotting ggplot2. FLCore and FLBRP can be installed from the FLR repository using the following from the R command line

```
install.packages(repos="http://flr-project.org/R")
```

While mixdist and gglot2 can be installed via the package menu in R.

More information on all these packages can be found at

http://flr-project.org

http://had.co.nz/ggplot2/

http://www.math.mcmaster.ca/peter/mix/mixdist.pdf

The data can be obtained from either the ICCAT website or the google code Project ???

```
#### R Packages
library(FLCore)   #FLR
library(FLBRP)    #Biological Reference Points
library(mixdist) #used for fitting mixture distributions
library(ggplot2) #plotting library

plusGroup=10

## growth data
vB       =c(Linf=238.6,K=0.185,t0=-1.404)

## Get the Catch-at-size data
# If read in from CSV file on http:\\www.iccat.int
#cas      =read.table("casSWOm8508_v2.csv",sep=";",header=T)[,c(3:7,9:10)]
#names(cas)=c("year","flag","fleet","gear","ld","len","n")

# If read on from R Data set file
load("cas.RData")

## adjust lengths to mid-year
cas$len   <-cas$len+0.5

## Plot CAS by decade
cas$yr    <-cas$year %% 10
cas$decade<-cas$year-cas$yr

p=ggplot(cas) +
    geom_histogram(aes(len,weight=n),colour="darkgreen",fill="white",binwidth=5) +
    scale_x_continuous(name="Length") + scale_y_continuous(name ="Frequency")   +
    facet_grid(yr~decade)
p
```
Initial Chunk:

ii) Functions

Several function were written to aid in the analysis, these were for the von Bertalanffy growth equation, an age slicing algorithm and a wrapper for the mix() method in mixdist. The latter as well as fitting the length distributions-at-age generated summary plots.

```
#### Growth
vonBert=function(age,Linf, K, t0) Linf*(1.0-exp(-K*(age-t0)))

#### Age Slicing
ageIt=function(len=NULL,n=NULL,Linf=NULL,K=NULL,t0=0.0,timing=0.5,plusGroup=30){
     ## expected age at length adjust to beginning of year
     age=pmax(pmin(floor(t0-log(1-pmin(len/Linf,.9999999))/K+timing),plusGroup),0)

     ## calculate frequencies
     res=aggregate(n, list(age=age), sum)

     return(res)}


#### Wrapper function for statistical ageing with a plot
cas2caa<-function(x,grwPar,constr,ages){
  ## aggregate frequencies by bins
  tst    =with(x,aggregate(n,list(len=len),sum))
  lnDist=mix(tst,grwPar,"norm",constr=constr,emsteps=3,print.level=1)

  plot(lnDist,xlim=c(50,250))
  abline(v=vonBert(ages,vB["Linf"],vB["K"],vB["t0"]),col="green",lty=5)
  mtext(unique(x$year))

  return(data.frame(age=ages,lnDist$parameters))}
```
Functions Chunk:

iii) Age Slicing Chunk:

Perform age slicing and create catch-at-age matrix.

```
#### use plyr package (also called by ggplot2) to process by year
caa=ddply(cas, .(year), function(x,vB,plusGroup,timing)
       ageIt(x$len,x$n,vB["Linf"],vB["K"],vB["t0"],timing,plusGroup), vB, plusGroup, 0.5)

#### Create an FLQuant for FLR to take advantage of fisheries stuff
names(caa)[3]="data"
caaSlice     =as.FLQuant(caa)
```

Age Slicing Chunk:

iv)   Statistical Ageing Chunk:

Statistical age estimation is performed using the mix() method. This requires initial guesses for the parameters to be estimated set using mixparam, which also allows constraints to be set. The initial guesses are the age slicing estimates; mean lengths-at-age are only estimated for the first three ages as there is little information to do so for older ages, likewise it is assumed that the CV is the same for all ages.

```
## constraints
constr<-mixconstr(conmu ="MFX", fixmu=c(rep(FALSE,3),rep(TRUE,8)), consigma ="CCV")

ages<-0:10

## Set initial proportions based on age slicing,
## expected length on Von Bertalanffy growth equation,
#
# sigma on expert knowledge!
grwPar=mixparam(pi    =c(apply(sweep(caaSlice,2,apply(caaSlice,2,sum),"/"),1,mean)),
                mu    =vonBert(ages+0.5,vB["Linf"],vB["K"],vB["t0"]),
                sigma =0.05) #ageCV)
grwPar$sigma=grwPar$sigma*grwPar$mu

## plot while fitting, again using plyr function ddply
par(mfrow=c(5,5), mar=c(1,1,1,1))
ageThem    =ddply(cas,.(year),cas2caa,grwPar=grwPar,constr=constr,ages=ages)

## process so you get an FLQuant
caaStat    =merge(ageThem,with(cas,aggregate(n,by=list(year=year),sum)))
caaStat$data=caaStat$x*caaStat$pi
caaStat    =as.FLQuant(caaStat[,c("age","year","data")])
```
Statistical Ageing Chunk:

v)   Catch Proportions Chunk:

Plot the catch-at-age proportions obtained from the two methods, scaled by maximum number-at-age within a year.

```
#### Catch proportions
## scale catch by max within a year put data into data.frame
dat<-rbind(data.frame(Method="Age Slicing",
                      as.data.frame(sweep(caaSlice,2,apply(caaSlice,2,max),"/"))),
           data.frame(Method="Statistical estimation",
                      as.data.frame(sweep(caaStat, 2,apply(caaStat, 2,max),"/"))))

## plot and compare
ggplot(dat) + geom_point(aes(year,age,size=abs(data))) +
              scale_area(to=c(3,10)) +
              facet_wrap(~Method)    +
              opts(legend.position = "none")
```
Catch Proportions Chunk:

vi) Catch Curve Chunk:

Compare the estimates of Z obtained from the two catch-at-age matrices, by performing a year class catch curve analysis.

```
#### Catch Curves
dat<-rbind(data.frame(Method="Age Slicing",
                      as.data.frame(FLCohort(log(caaSlice)))),
           data.frame(Method="Statistical estimation",
                      as.data.frame(FLCohort(log(caaStat )))))

ggplot(dat) + geom_point(aes(age,data,group=cohort,col=cohort)) +
          geom_smooth(aes(age,data)) +
          facet_wrap(~Method) +
          scale_y_continuous(limits=c(0,12))
```
Catch Curve Chunk:

vii) Catch Ratio Chunk:

Compare the estimates of Z obtained from the two catch-at-age matrices, by calculating the log catch ratios

```
#### function to calculate log catch ratios down a cohort
z=function(x) FLCohort(log(x[1:(dim(x)[1]-1),-dim(x)[2]]/x[2:(dim(x)[1]),-1]))

## get data in data.frame for gglot
dat<-rbind(data.frame(Method="Age Slicing",
                      as.data.frame(FLCohort(z(caaSlice)))),
           data.frame(Method="Statistical estimation",
                      as.data.frame(FLCohort(z(caaStat )))))

## plotting
p = ggplot(dat)

p + geom_line(aes(age,data,col=cohort,group=cohort))     +
    stat_smooth(aes(age,data), method="loess",span=0.3) +
    facet_wrap(~Method)

p + geom_line(aes(cohort+age,data,col=age,group=age)) +
    stat_smooth(aes(cohort+age,data), method="loess",span=0.3) +
    facet_wrap(~Method) +
    scale_x_continuous(limit=c(1985,2005))
```
Catch Ratio Chunk:

viii) Reference Points Chunk:

Compare reference points based upon selection patterns derived from the two methods, based upon 2010 assessment.

```
## 2010 assessment
load("swo2009PG10.RData")
swoBrp<-FLBRP(swo2009PG10,nyears=21)
fbar(swoBrp)<-apply(fbar.obs(swoBrp),1,mean)

ggplot(as.data.frame(harvest(swo2009PG10))) +
      geom_point(aes(age,data))+stat_smooth(aes(age,data))

objFn<-function(par,x,y){
  landings.sel(x)[]<-par
  ctch<-catch.n(brp(x))
  ctch<-sweep(ctch,2,apply(ctch,2,max),"/")
  y   <-apply(sweep(  y,2,apply(  y,2,max),"/"),1,mean)

  return(log(sum((ctch-y)^2)))}

par<-c(catch.sel(swoBrp))
objFn(par,swoBrp,caaStat)
res<-optim(par,objFn,x=swoBrp,y=caaStat)

swoBrp2<-swoBrp
landings.sel(swoBrp2)[]<-res$par
fbar(swoBrp) <-FLQuant(seq(0,1,length.out=51))
fbar(swoBrp2)<-FLQuant(seq(0,1,length.out=51))

swoBrp <-brp(swoBrp)
swoBrp2<-brp(swoBrp2)


rpts<-data.frame(rbind(refpts(swoBrp )[1:2,1:5,1,drop=T],
                       refpts(swoBrp2)[1:2,1:5,1,drop=T]))
rpts<-cbind(slicing=rep(c("deterministic","stochastic"),each=2),
            refpt  =rep(c("F0.1","FMSY"),2),rpts)

ts<-rbind(cbind(slicing="deterministic",model.frame(FLQuants(Fbar=fbar(swoBrp),
SSB=ssb(swoBrp), Yield=yield(swoBrp )))[,c("year","Fbar","SSB","Yield")]),
          cbind(slicing="stochastic",
model.frame(FLQuants(Fbar=fbar(swoBrp2),SSB=ssb(swoBrp2),Yield=yield(swoBrp2)))[,c("year","F
bar","SSB","Yield")]))


ggplot(ts)+
  geom_line(aes(Fbar,Yield,col=slicing))+
  geom_point(data=rpts,aes(harvest,yield,col=slicing),size=4,pch=19)
```

Reference Points Chunk:

ix) Monte Carlo Chunk:

Estimate uncertainty in age slicing estimates using Monte Carlo simulation.

```
#### Age Slicing
## function
mcAgeSlice<-function(x,vB,timing,plusGroup) {
   x<-x[sample(1:dim(x)[1],dim(x)[1],TRUE),]
   ageIt(x$len,x$n,vB["Linf"],vB["K"],vB["t0"],0.5,10)[,2]}

cas1985<-cas[cas$year==1985,]
cas2009<-cas[cas$year==2009,]
mcAS<-array(NA,c(2,11,1000))
for (i in 1:1000) mcAS[1,,i]<-mcAgeSlice(cas1985, vB, timing, plusGroup)[,2]
for (i in 1:1000) mcAS[2,,i]<-mcAgeSlice(cas2009, vB, timing, plusGroup)[,2]

## CVs
apply(mcAS,1:2,function(x) (var(x)^0.5)/mean(x))

#### Statistical estimates
tst   =with(cas[cas$year==2009,],aggregate(n,list(len=len),sum))
lnDist=mix(tst,grwPar,"norm",constr=constr,emsteps=3,print.level=1)
lnDist$se/lnDist$parameters

Monte Carlo  Chunk:
```