0.1 The package a4adiags

The package 'a4adiags' contains some additional diagnostics based on the reference.Runs test checks weather the residuals are randomly distributed. A "run" is a sequence of the same sign residuals. Few runs indicate a trend or a correlation in the residuals while too many runs may suggest overfitting. The primary output of a runstest is a p-value where: - A high p value ($p \leq 0.05$) suggests that the residuals are randomly distributed - A low p value indicates a non-random pattern in the residuals

```
library(FLa4a)

## Loading required package:  FLCore
## Loading required package:  lattice
## Loading required package:  iterators
## FLCore (Version 2.6.20.9201, packaged:  2024-07-08 21:29:30 UTC)
## Loading required package:  triangle
## This is FLa4a 1.9.0.  For overview type 'help(package="FLa4a")'

library(FLCore)
library(a4adiags)

## Loading required package:  ggplot2
##
## Attaching package:  'ggplot2'
## The following object is masked from 'package:FLCore':
##
##      %+%
## Loading required package:  ggplotFL

library(ggplot2);theme_set(theme_bw())
rm(list = ls())
fit05 <- sca(mut09, mut09.idx, fmod = ~factor(age) + s(year, k = 8))
res05 <- residuals(fit05, mut09, mut09.idx)
```

```
plotRunstest(fit05, mut09.idx, combine = F) + theme_bw() + facet_wrap(~age)
```

```
plotRunstest(catch.n(mut09), catch.n(mut09 + fit05), combine = F) + theme_bw() + facet_wrap(
```

Green shading indicates no evidence ($p < 0.05$) and red shading evidence ($p > 0.05$) to reject the hypothesis of a randomly distributed time-series of residuals, respectively. The shaded (green/red) area spans three residual standard deviations to either side from zero, and the red points outside of the shading violate the ' 3σ limit' for that series.

Hindcast is used to assess the prediction skill of the model by removing a couple of years from the end of the time series and forecasting for that years, to
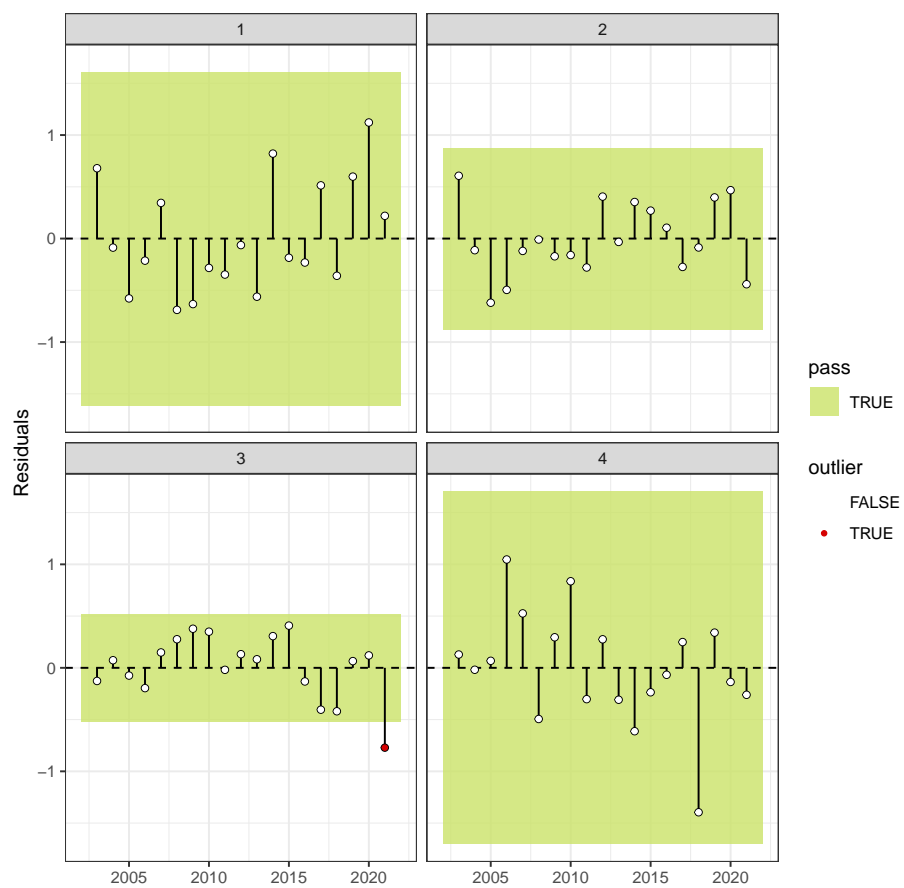
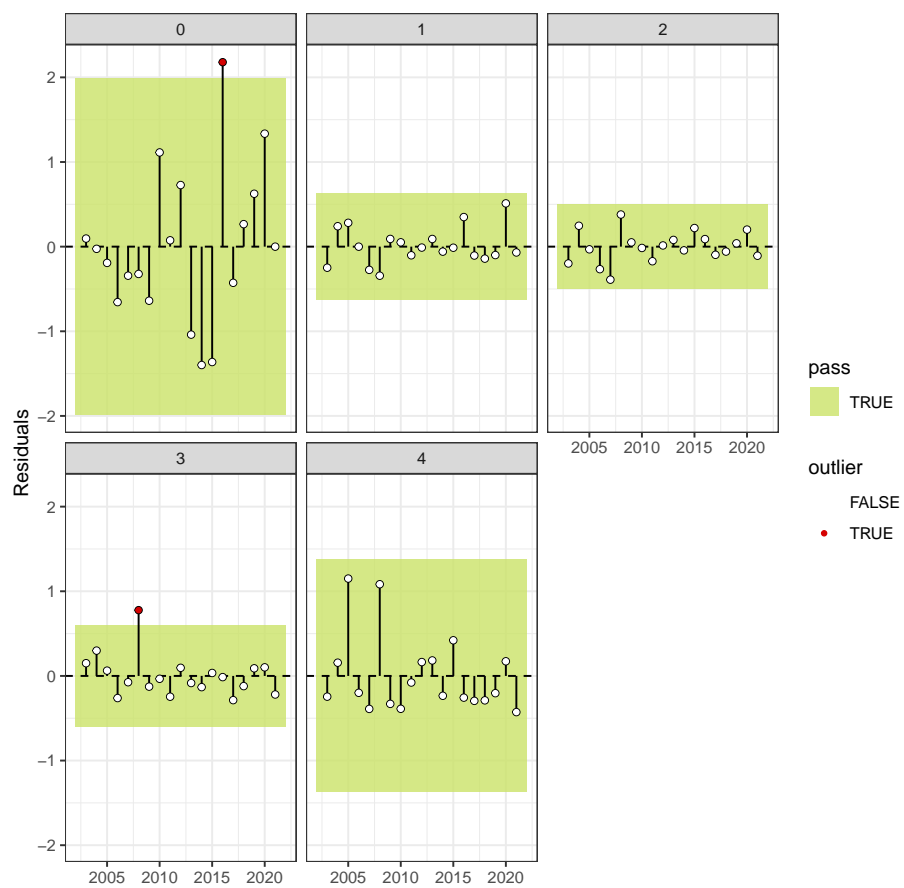Figure 1: Runstest for the index

Figure 2: Runstest for the catch

assess how close the projected values are to the ones assessed by the model. The hindcast can estimate forecast bias by comparing the forecasted values to the reference model estimates. The 'a4adiags' package use a different approach, it is based on a technique proposed by Kell ref named hindcast cross-validation where the forecasted values are compared with the observed ones. Additionally mean absolute squared error is computed, a statistic for evaluating the prediction skill. MASE basically compares the model prediction skill against a random walk, i.e against the predicted value of a random process based only on the previous year's observation.

```
xval <- a4ahcxval(mut09, FLIndices(mut09.idx), nyears = 5, nsq = 3, fmodel =   ~factor(age) +

## Warning:   executing %dopar% sequentially:   no parallel backend registered

plotXval2(xval$indices) + ggtitle(paste0("Hindcast"))
```
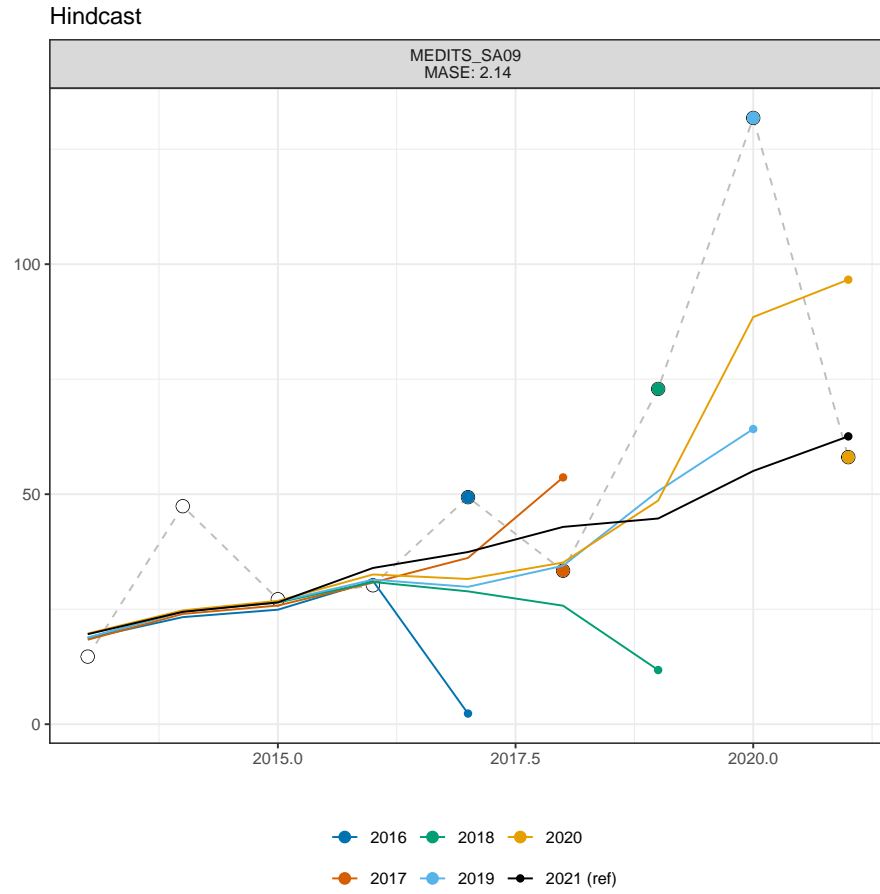
Figure 3: Hindcasting and MASE statistic. A MASE score ¿ 1 indicates that the average model forecasts are worse than a random walk. Conversely, a MASE score of 0.5 indicates that the model forecasts twice as accurately as a naïve baseline prediction; thus, the model has prediction skill