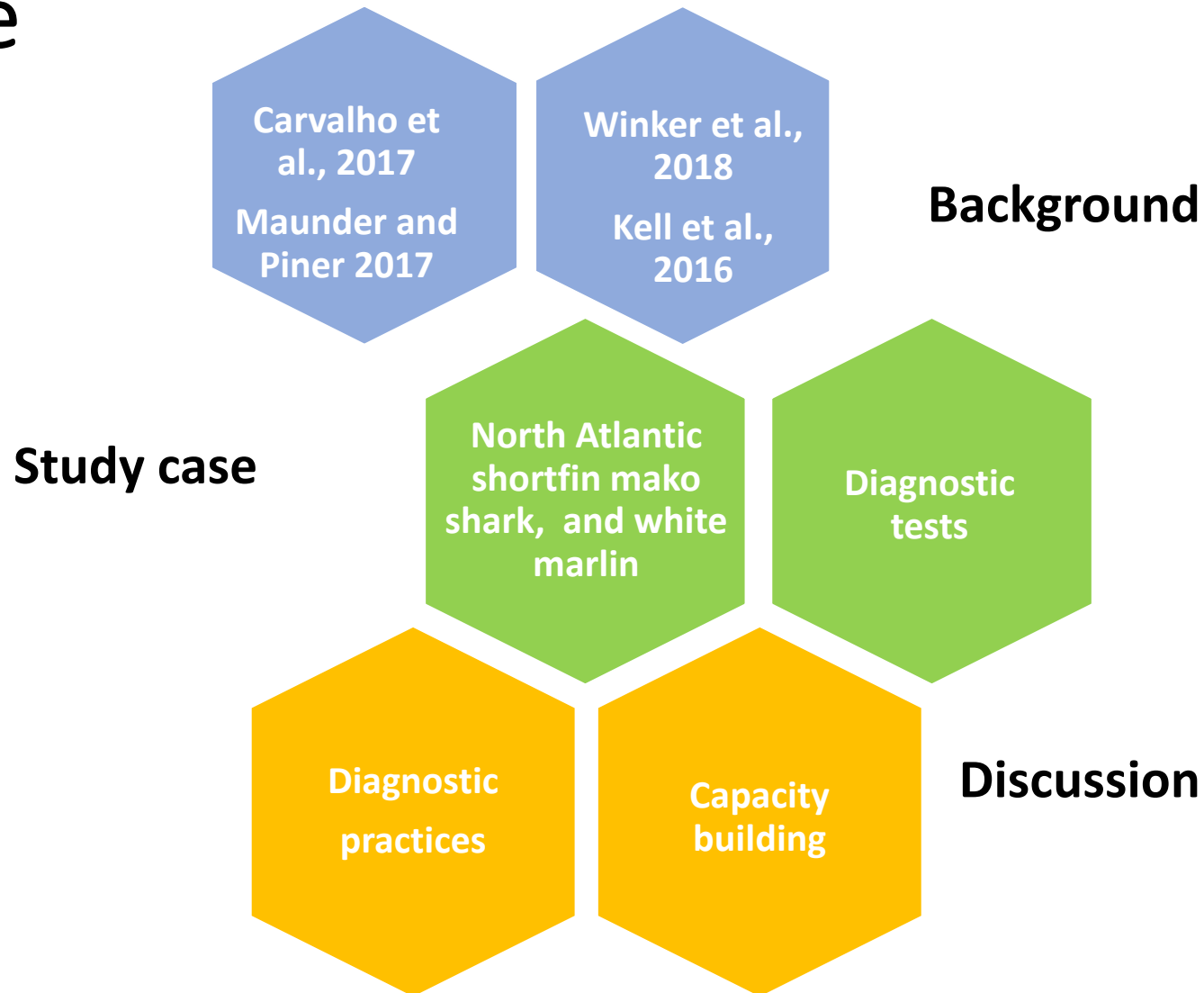# A Cookbook for Using Model Diagnostics in Integrated Stock Assessments

Felipe Carvalho, Henning Winker, Laurence Kell, **Dean Courtney**, Massimiliano Cardinale, Dawit Y. Ghebrehiwet, Michael Schirripa, Maia Kapur, Kevin R. Piner, and Mark N. Maunder

CAPAM workshop on the creation of frameworks for the next generation general stock assessment models

Outline

Carvalho et al., 2017
Maunder and Piner 2017

Winker et al., 2018
Kell et al., 2016

Background

Study case

North Atlantic shortfin mako shark, and white marlin

Diagnostic tests

Diagnostic practices

Capacity building

Discussion

Background

Carvalho et al., 2017

Maunder and Piner 2017

Winker et al., 2018

Kell et al., 2016

Full length article

Can diagnostic tests help identify model misspecification in integrated stock assessments?

CrossMark

Felipe Carvalho [a,b,*], André E. Punt [c], Yi-Jay Chang [d], Mark N. Maunder [e,f], Kevin R. Piner [g]

- Diagnostics that identify misfit to data or conflict among model fits to different data components are important tools to identify potential misspecification in integrated models

- Carvalho et al (2017) tested several diagnostics (residual analysis, retrospective analysis, R0 likelihood component profile, and age-structured production model - ASPM, among others) on simulated data sets with imposed misspecifications

- No single diagnostic worked well in all of the cases they evaluated with simulation

- They recommended the use of a carefully selected range of diagnostics that proved to increase the ability to detect model misspecification in their simulated data sets

Full length article

Dealing with data conflicts in statistical inference of population assessment models that integrate information from multiple diverse data sets

CrossMark

Mark N. Maunder [a,b,*], Kevin R. Piner [c]

- Maunder and Piner (2017) developed an algorithm based on diagnostic tests (including residual analysis, R0 likelihood profile, and ASPM) to guide the construction of stock assessment models and reduce model misspecification during the model construction process

- Their algorithm for model construction, emphasized two components; avoiding data conflicts (or facilitating their interpretation) and diagnosing and fixing data conflicts

- They also provided a flowchart diagram to help users to complete the various steps involved in the model construction in the correct order and they evaluated the algorithm on simulated data sets with imposed misspecifications that were not disclosed to the analyst
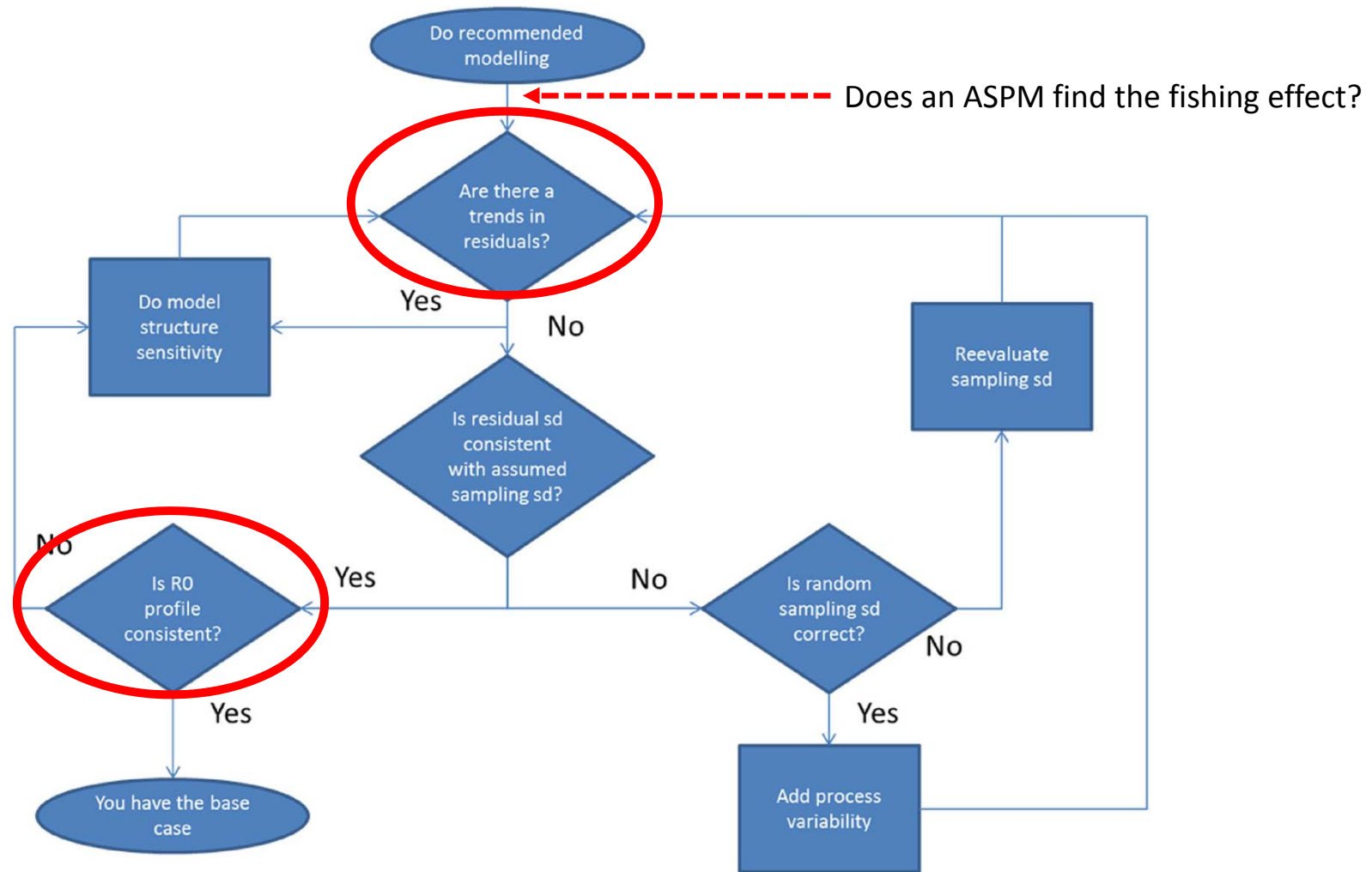
Flow diagram from Maunder and Piner (2017)

JABBA: Just Another Bayesian Biomass Assessment

Henning Winker[a,b,*], Felipe Carvalho[c], Maia Kapur[c,d]

JABBA residual plot was adapted for integrated age-structured models

Evaluation of the prediction skill of stock assessment using hindcasting

CrossMark

Laurence T. Kell[a,*], Ai Kimoto[b], Toshihide Kitakado[c]

- Prediction skill diagnostics (hindcasting) utilizing the Fisheries Library in R (FLR)…

- Recent example of application in IOTC

IOTC-2019-WPTT21-48

Alternative Assessment for the Indian Ocean Yellowfin Tuna Stock; with Generic Goodness of Fit Diagnostics.

Laurence Kell and Rishi Sharma

Our Diagnostics Approach – After Base Case Model Developed
Use a Carefully Selected Range of Diagnostics

Study Case

North Atlantic shortfin mako shark, and white marlin

Diagnostic tests
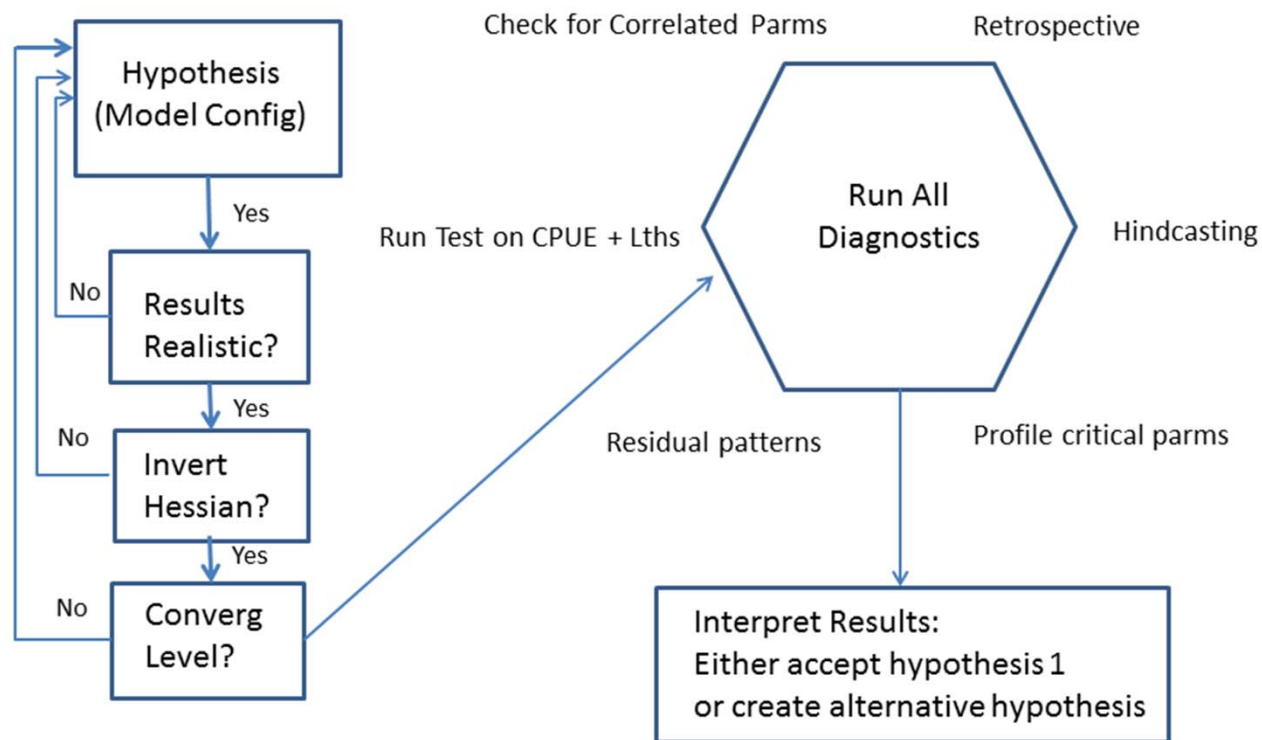
Courtney, D., Carvalho, F., Winker, H., and L. Kell. 2019. Examples of Stock Synthesis diagnostic methods and results implemented for previously completed North Atlantic shortfin mako Stock Synthesis model runs. SCRS/2019/088 (in press) .

Schirripa, M. J. 2019. Current status of the white marlin (*Kajikia albida*) stock in the Atlantic Ocean 2019: predecisional stock assessment model. Document SCRS/2019/110 (in press): 28 pp
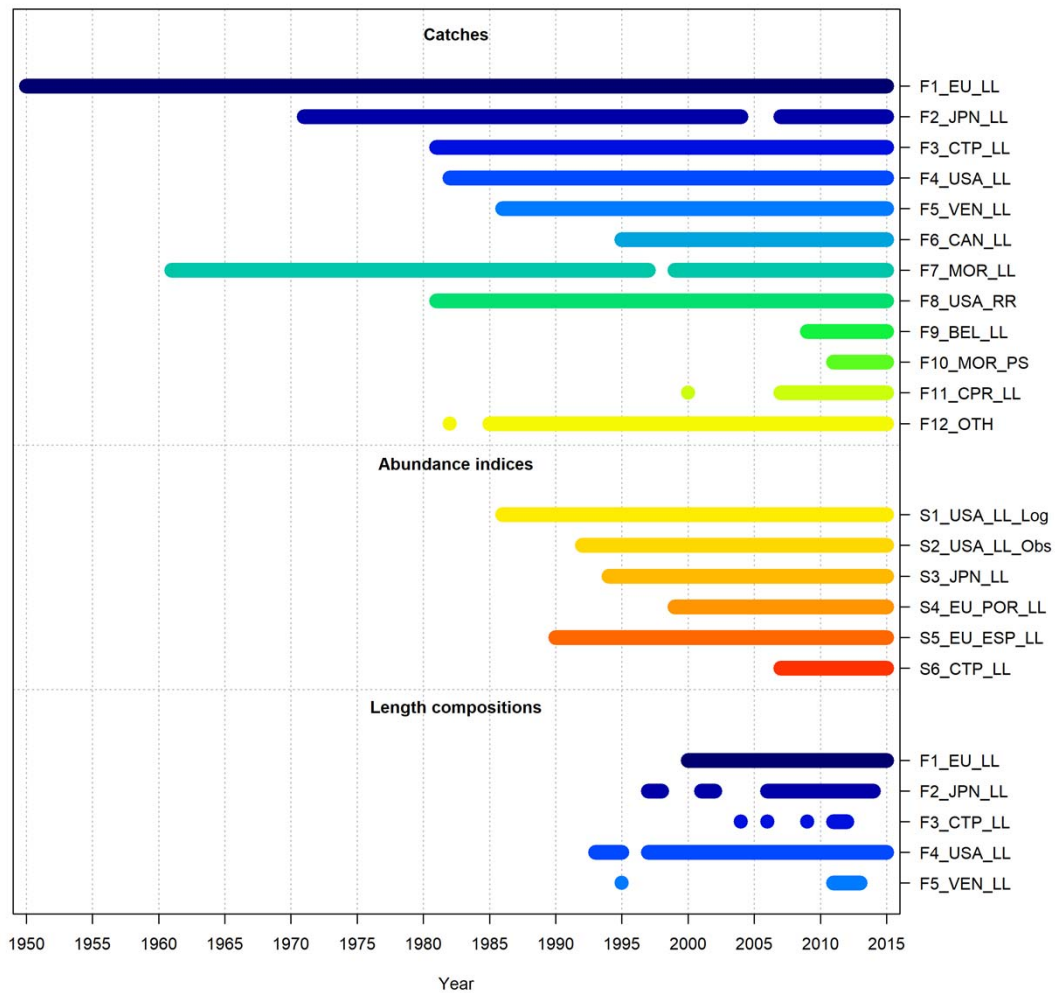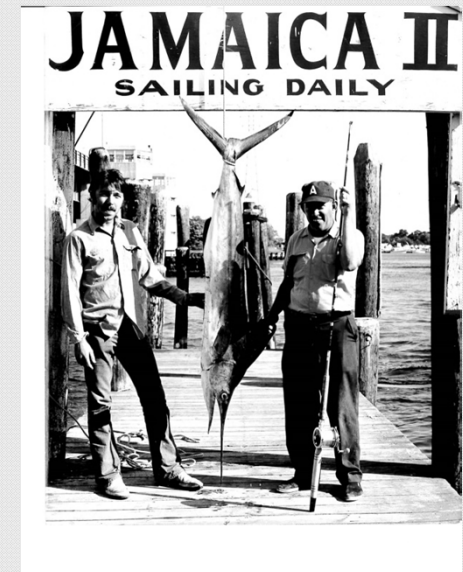
SHORTFIN MAKO STOCK IN THE NORTH ATLANTIC OCEAN
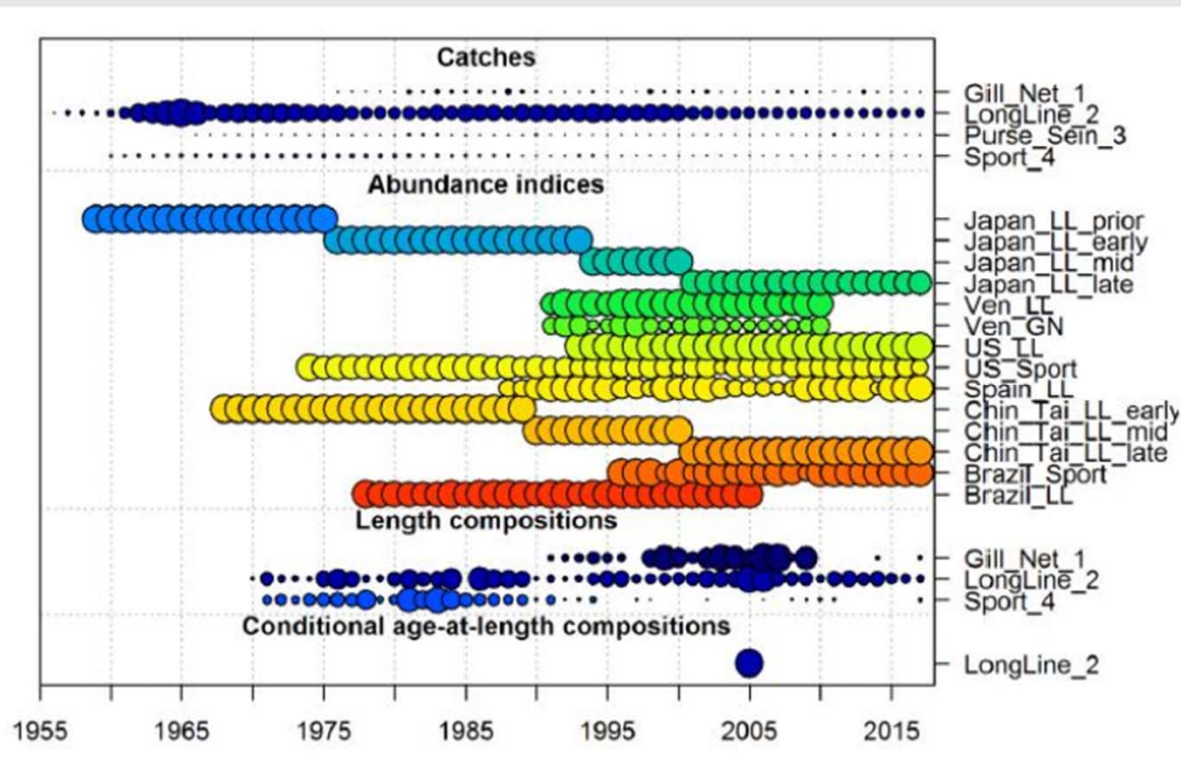
WHITE MARLIN (*Kajikia albida*) STOCK IN THE ATLANTIC OCEAN

Model Diagnostics Evaluated
After Model Convergence

- Model convergence was based on whether or not the Hessian inverted (i.e., the matrix of second derivatives of the likelihood with respect to the parameters, from which the asymptotic standard error of the parameter estimates is derived).
- Other convergence diagnostics were also evaluated.
    - Excessive CVs on estimated quantities ($\gg$ 50%) or a large final gradient (>1.00E-05) were indicative of uncertainty in parameter estimates or assumed model structure.
    - The correlation matrix was also examined for highly correlated (> 0.95) and non-informative (< 0.01) parameters.
    - Parameters estimated at a bound were a diagnostic for possible problems with data or the assumed model structure.

Diagnostic 1* - Jittering

- Jittering the starting values of the parameters to evaluate whether the model converges to a global solution, rather than a local minimum.

- 200 iterations of the jitter test for global convergence resulted in 56 model runs with the a minimum total likelihood value equal to that of the base case model run (198.5 likelihood units), two model runs with higher total likelihood values (247.7 and 268.7 likelihood units, respectively), and 142 model runs that failed to converge.
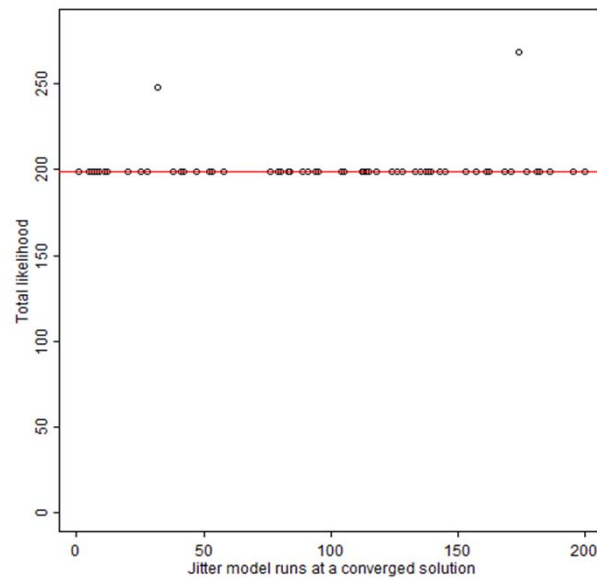
Convergence

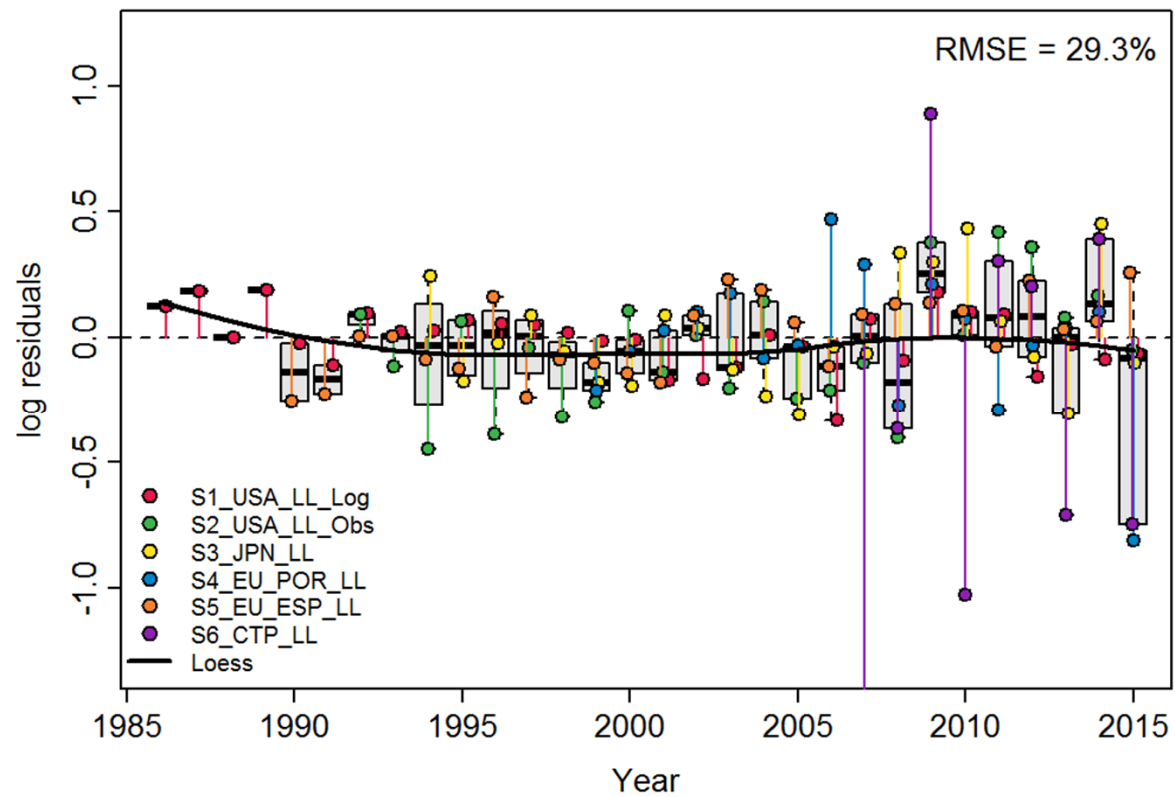## Diagnostic 2* - MCMC Diagnostics

- MCMC diagnostics for each model run were evaluated with both a relatively short and a relatively long chain.

- Convergence of the MCMC samples to the posterior distribution was evaluated here with a visual inspection of the trace along with 'Heidelberger and Welch' and 'Geweke' tests implemented in the coda package.
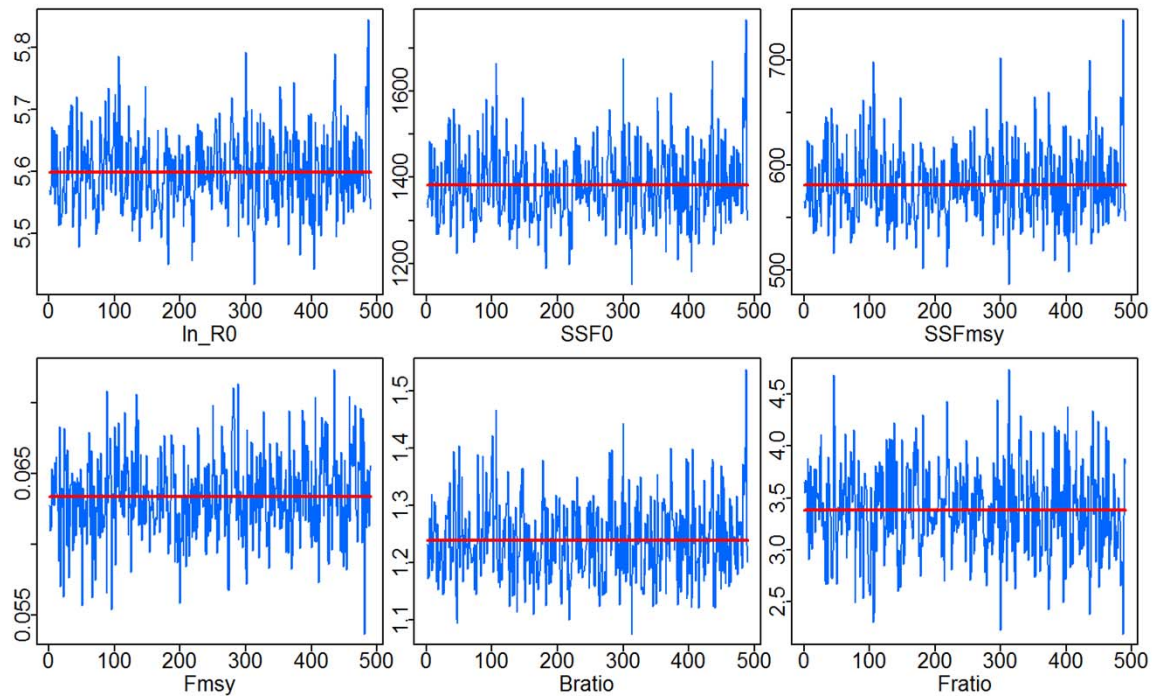

*Convergence

Diagnostic 3 - Residual Patterns

- JABBA-residual Plot (CPUE Residuals) fit a smoother to log scale residuals of all CPUE indices fit in model.

- Adapted from JABBA (Winker et al. 2018) for Stock Synthesis and implemented in R.
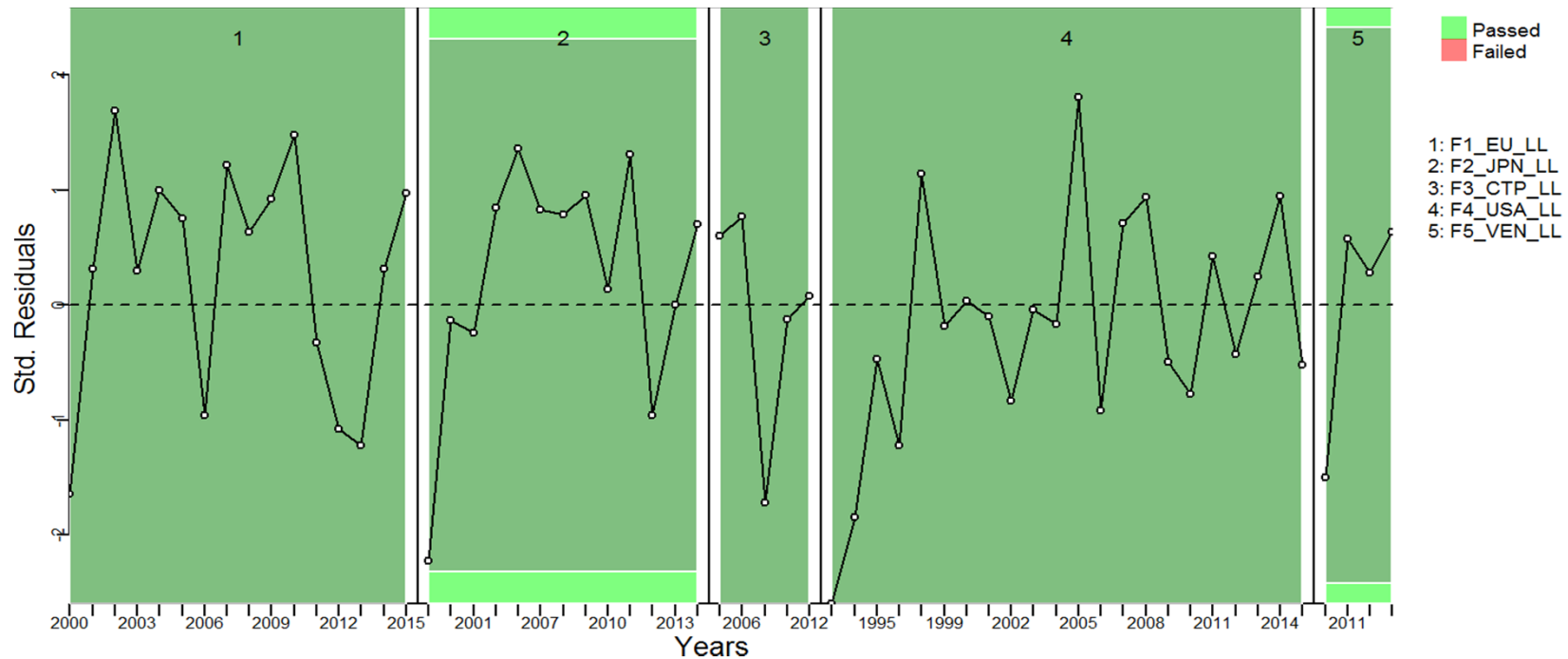
Diagnostic 3 - **Residual Patterns (Randomness)**

- A runs test was applied to the residuals of each CPUE index fit in the Stock Synthesis model in order to quantitatively evaluate the randomness of the time-series of CPUE residuals by fleet.

- R plots were developed to visualize results obtained from residuals runs tests

- Individual time-series data points further than three standard deviations away from the mean (the three-sigma rule), which is another test used to detect non-random time series (e.g., see Anhøj and Olesen 2014)
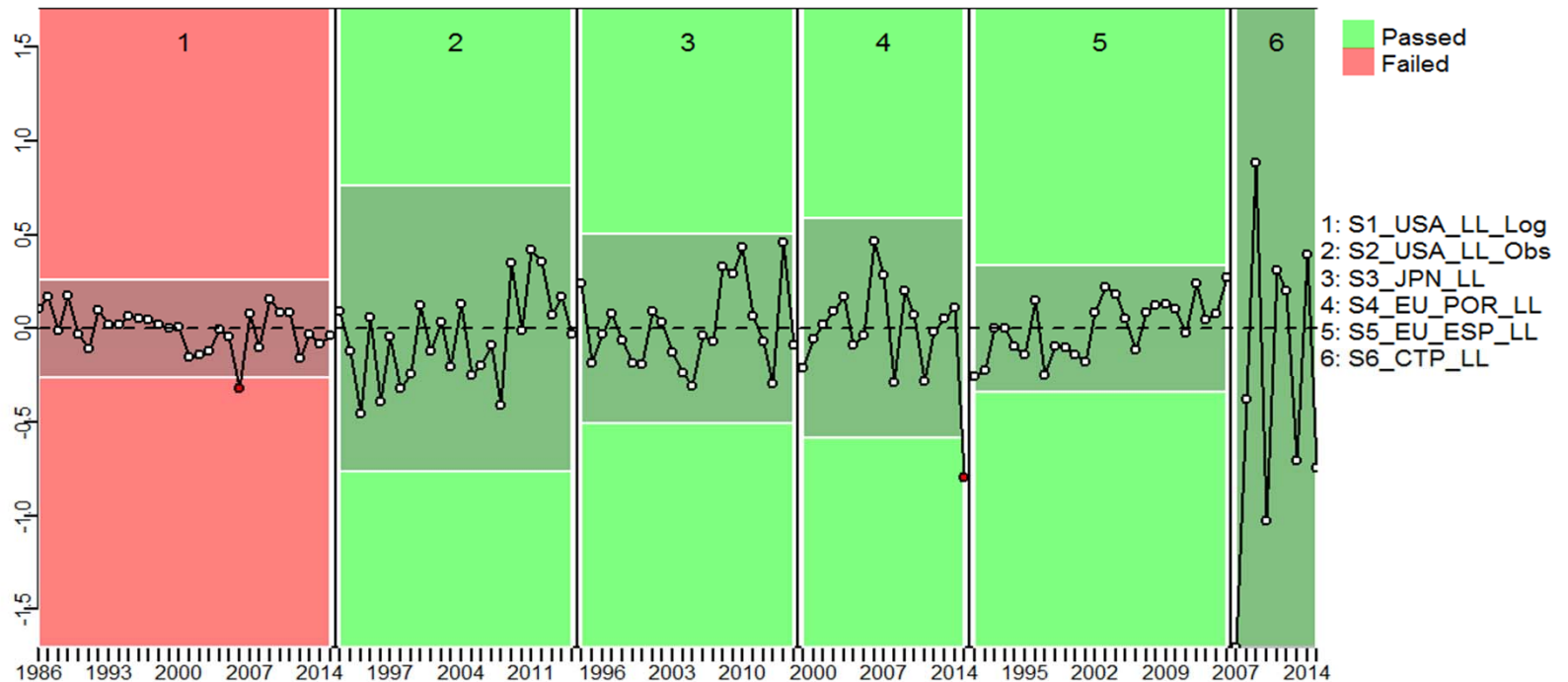
Diagnostic 3 - Residual Patterns (Randomness)

- The runs test was also applied to the standardized residuals of the fit to length composition by fleet and year in order to quantitatively evaluate the randomness of the time-series of length composition residuals by fleet

- Standardized residuals were obtained for each fleet using the Francis method (Carvalho et al. 2017, citing Punt 2017 their Table 2 equation 1.C; e.g., see Francis 2011, 2014, 2017)

## *Diagnostic 3* - **Residual Patterns (Randomness)**

- The runs test was also applied to recruitment deviations estimated in the Stock Synthesis model in order to quantitatively evaluate the randomness of the time-series of estimated recruitment deviations.
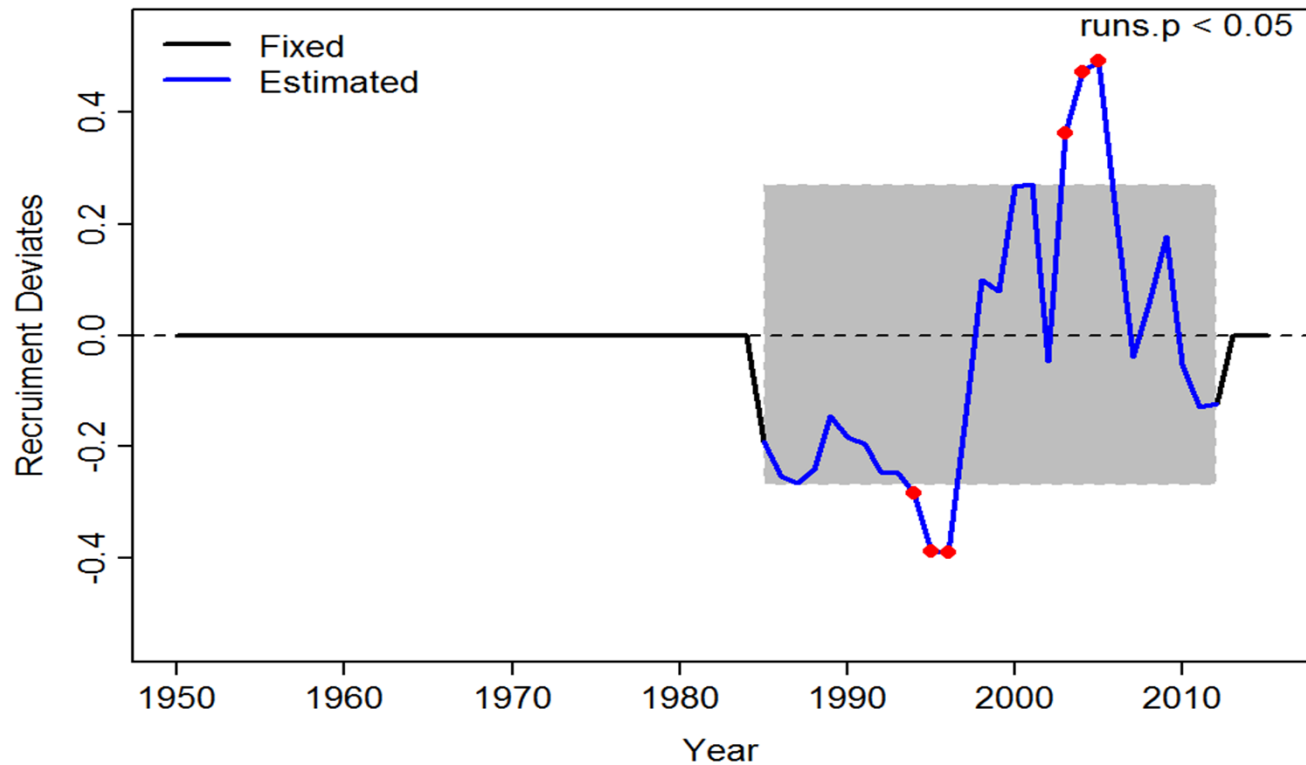
- Implemented using the function 'runs.test' in the R package 'tseries' (Trapletti, 2011); a nonparametric randomness hypothesis test.

Diagnostic 4 - **Age-structured Production Model Diagnostic (ASPM)**

- An age-structured production model diagnostic (ASPM; e.g., Maunder and Piner 2017, Carvalho et al. 2017) was applied to the Stock Synthesis model results

- The models showed similar overall trend, however after the 1990's the ASPM showed a less steep decline in spawning stock size than the full integrated stock assessment model; The asymptotic 95% confidence intervals of relative spawning stock size did not overlap for many of the most recent years
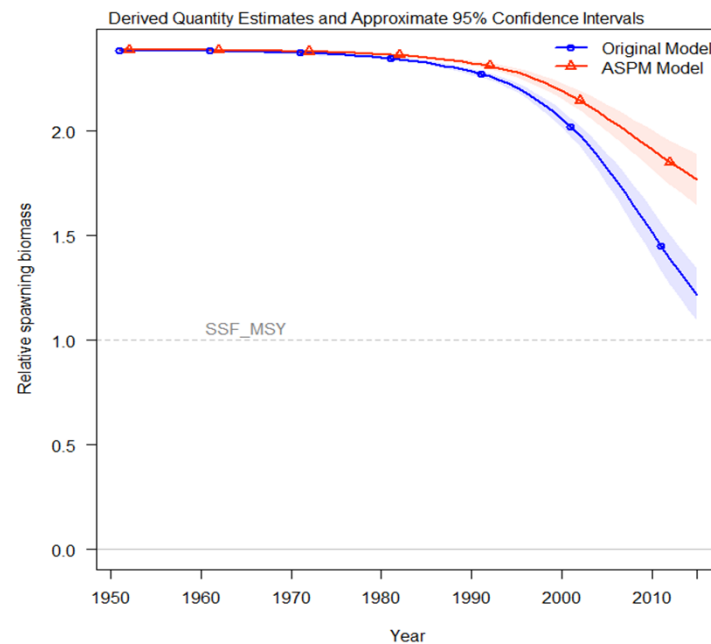
Diagnostic 5 - R0 Likelihood Component Profile

- An R_0 likelihood component profile (e.g., Carvalho et al. 2017) was applied to the results; The diagnostic was implemented here by sequentially fixing the equilibrium recruitment parameter, R_0, on the natural log scale, $\log(R_0)$
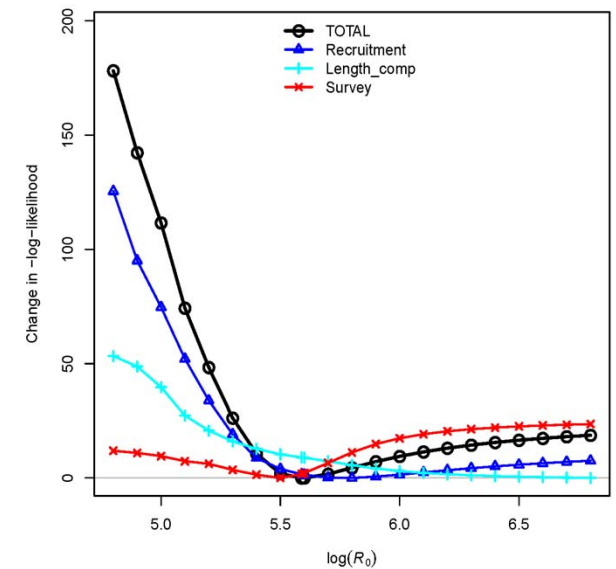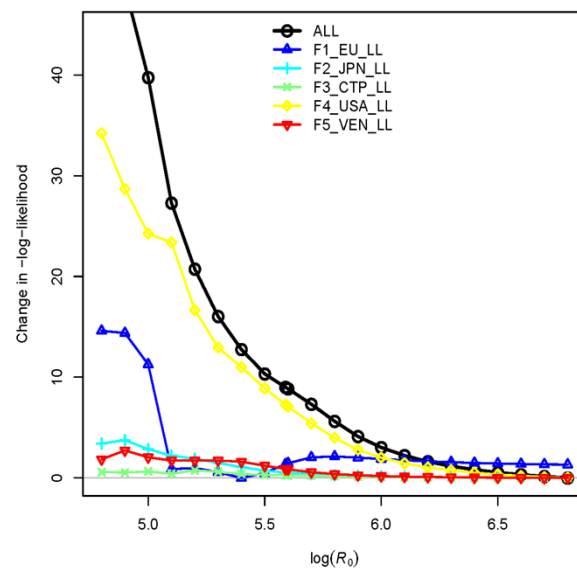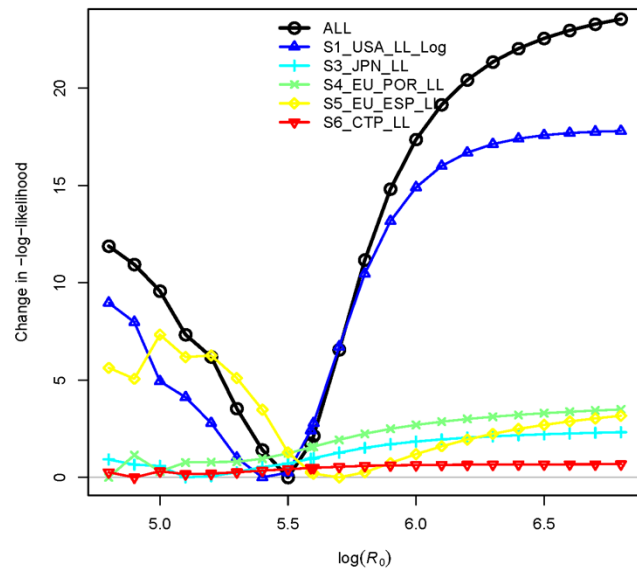
Diagnostic 6 - **Retrospective Analysis**

- Retrospective analysis is a way to detect bias and model misspecification (e.g., Hurtado-Ferro et al. 2014); The diagnostic was implemented here by sequentially eliminating the five most recent years of data from the full stock assessment model

Derived Quantity Estimates and Approximate 95% Confidence Intervals

Mohn's rho = 0.064
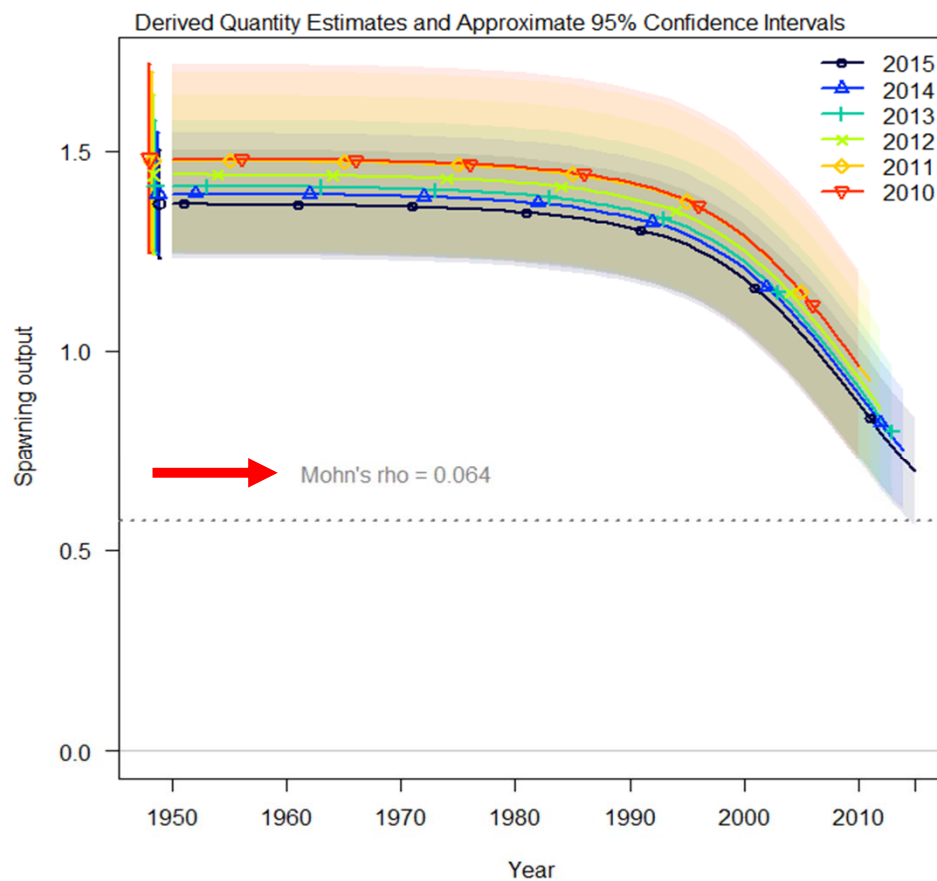
Diagnostic 7 - Prediction Skill (Hindcast Precision)

- In addition to determining if the model fits the historical data, it is important to evaluate whether the model can replicate the future dynamics of the system, which is required to provide management advice

- One diagnostic for this is model prediction skill; Model prediction skill was diagnosed here with hindcasting precision (Kell et al. 2016), an extension of Retrospective Analysis

- Using a hindcast, each assessment model was retrospectively re-run by tail cutting, i.e. removing recent years' data and the biomass trajectories projected up to the most recent year

- Model-free validation was adapted here for Stock Synthesis to compare the observed CPUE indices in the recent years (the input data) to their out of sample predicted values (the hindcast) calculated by multiplication of catchability and vulnerable biomass obtained from the stock assessment model one-step ahead predicted values from each hindcast for up to 15 years

- Hindcast results were summarized using the Mean Absolute Scaled Error (MASE); A scaled error is less than one if it arises from a better forecast than the average one-step-ahead naïve forecast (equal to the last observation).

Diagnostic 7 - **Prediction Skill Continued**

- MASE scores for the CPUE indices EU_ESP_LL, and JPN_LL were greater than one. This diagnostic result indicated that the average one-step ahead naïve forecast was a better predictor than the stock assessment model for those indices, i.e. knowledge of resource dynamics in these cases did not help in prediction of those indices.
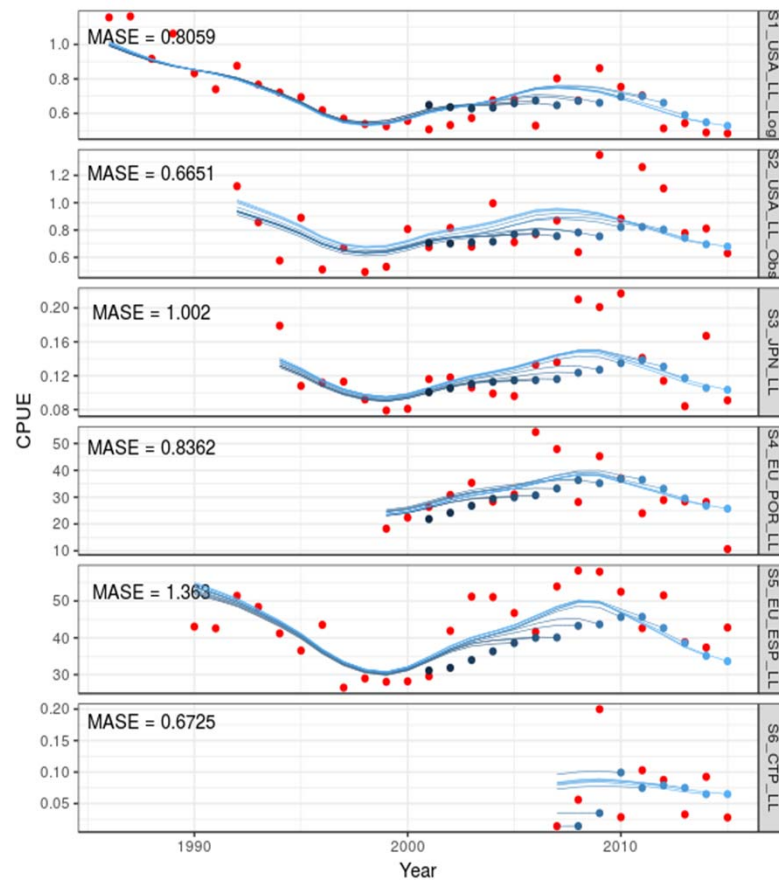
Case Study Results

Diagnostic and prediction skill results were consistent

Significant non-random Rec devs in combination with failed ASPM and poor out of sample prediction skill may indicate miss-specification of the system model production function (Maunder and Piner 2017)

Next Steps:
Evaluate model for miss-specification

E.g., Following flow chart in Maunder and Piner (2017)

Perform sensitivity analyses to assumed steepness and natural mort

Develop alternative model hypotheses

Rerun all model diagnostics

Repeat…

Discussion

Diagnostics practices

Capacity building

Focus Questions Addressed

- Coding philosophies and software structure
  - Is there a way to easily allow the addition of new features?

- Stock assessment model features
  - What system should be used to decide what features are included in the next generation model?

- User interface and good practices defaults
  - What kinds of comprehensive user interfaces are worth the effort?

Software

Software (and versions) used for each diagnostic.

Diagnostic	Stock	R[1] and FLR[2]	r4ss[3]	Stock Synthesis[4]
JABBA-residual plot	Shortfin mako	R (version 3.3.3)	version 1.24.0	(version 3.24U)
Runs test of residuals	Shortfin mako	R (version 3.3.3)	version 1.24.0	(version 3.24U)
Retrospective analysis	Shortfin mako	R (version 3.4.4)	version 1.30.2	(version 3.24U)
R_0 likelihood component profile	Shortfin mako	R (version 3.4.4)	version 1.30.2	(version 3.24U)
Jitter	White marlin	R (version 3.4.4)	version 1.35.3	(version 3.30)
		Manually implemented		
ASPM	Shortfin mako	R (version 3.4.4)	version 1.30.2	(version 3.24U)
Prediction skill (hindcast)	Shortfin mako	FLR		

[1] R (R Core Team 2018). Available: https://www.R-project.org (May 2019).

[2] Fisheries Library in R (FLR; Kell et al. 2007). Available: http://www.flr-project.org (May 2019).

[3] R code for Stock Synthesis (r4ss; Taylor et al. 2018). Available: https://github.com/r4ss/r4ss (May 2019).

[4] Stock Synthesis (Methot and Wetzel 2013). Available: https://vlab.ncep.noaa.gov/web/stock-synthesis/home (May 2019).

Stock Synthesis

r4ss

- **Retrospective analysis**
  - SS_doRetro(); Ssgetoutput(); Sssummarize(); SSplotRetroRecruits(); Ssmohnsrho()

- **R_0 likelihood component profile**
  - SS_profile(); SSplotProfile; PinerPlot()

- **Jittering**
  - SS_RunJitter(); Ssgetoutput(); Sssummarize(); SSplotComparisons()

```R
# Retrospective analysis with Stock Synthesis
# Example application
# 2018 NOAA PIFSC Uku stock assessment
# Stock Synthesis (tested in version 3_30_05 for Windows)
# r4ss (tested in version 1.34.0)
# R (tested in version 3.5.3 64 bit)

library(r4ss)
#devtools::install_github('r4ss/r4ss')
#?r4ss

##################################################### Step 1
# Identify restrospective period
# e.g., for end.yr.vec    <- c(2015,2014,2013,2012,2011,2010)
start.retro <- 0      #end year of model e.g., 2015
end.retro  <- 5       #number of years for retrospective e.g., 2014,2013,2012,2011,2010

# Identify the base directory
dirname.base <- "C:\\Users\\felip\\Documents\\Uku"
dirname.base

# Identify the directory where a completed model run is located
dirname.completed.model.run <- paste0(dirname.base,'/retro')
dirname.completed.model.run

# Create a subdirectory for the Retrospectives
dirname.Retrospective <- paste0(dirname.base,'/Retrospective')
dir.create(path=dirname.Retrospective, showWarnings = TRUE, recursive = TRUE)
setwd(dirname.Retrospective)
getwd()

# Create a subdirectory for the Plots
dirname.plots <- paste0(dirname.Retrospective,"/plots_1")
dir.create(dirname.plots)

##################################################### Step 2
# Copy model files from the base run
file.copy(paste(dirname.completed.model.run,      "starter.ss_new", sep="/"),
          paste(dirname.Retrospective, "starter.ss", sep="/"))
file.copy(paste(dirname.completed.model.run,      "control.ss_new", sep="/"),
          paste(dirname.Retrospective, "CONTROL.SS", sep="/"))
file.copy(paste(dirname.completed.model.run,      "data.ss_new", sep="/"),
          paste(dirname.Retrospective, "DATA.SS", sep="/"))
file.copy(paste(dirname.completed.model.run,      "forecast.ss", sep="/"),
          paste(dirname.Retrospective, "forecast.ss", sep="/"))
file.copy(paste(dirname.completed.model.run,      "SS.exe", sep="/"),
          paste(dirname.Retrospective, "SS.exe", sep="/"))

##################################################### Step 3
starter <- readLines(paste(dirname.Retrospective, "/starter.ss", sep=""))

# 1) Starter File changes to speed up model runs
# Run Display Detail
#[8] "2 # run display detail (0,1,2)"
linen <- grep("# run display detail", starter)
starter[linen] <- paste0( 1 , " # run display detail (0,1,2)" )

# 2) Starter File changes made to be consistent with projections
linen <- grep("# Depletion basis:", starter)
starter[linen] <- paste0( 2 , " # Depletion basis:  denom is: 0=skip; 1=rel X*B0; 2=rel SPBmsy; 3=rel X*B_styr" )
linen <- grep("Depletion denominator", starter)
starter[linen] <- paste0( 1, " # Fraction (x) for Depletion denominator (e.g. 0.4)")
```

---

aputils

🕙 53 commits	⑂ 1 branch	⬡ 0 releases	👥 1 contributor

Branch: master ▾   New pull request   Create new file   Upload files   Find File   **Clone or download ▾**

👤 mkapur random   Latest commit 71a25b6 9 days ago

📁 .Rproj.user	random	9 days ago
📁 R	add simple kobe with ISC options	9 days ago
📁 man	why did these all get updated into man?	15 days ago
📁 tests	initial dump	a year ago
📄 .Rbuildignore	update 10 12	7 months ago
📄 .Rhistory	random	9 days ago
📄 .gitignore	ignore rhist	26 days ago
📄 DESCRIPTION	why did these all get updated into man?	15 days ago
📄 NAMESPACE	why did these all get updated into man?	15 days ago
📄 README.md	Update README.md	11 months ago
📄 kaputils.Rproj	initial dump	a year ago

📖 **README.md**

kaputils

custom plotting and simulation utilities

```
devtools::install_github("mkapur/kaputils")
library(kaputils)
```

Mahalo!