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
North Atlantic shortfin mako shark, and white marlin Capacity building Diagnostic practices Discussion Diagnostic tests Study case Background

- Carvalho et al., 2017
- Maunder and Piner 2017
- Winker et al., 2018
- Kell et al., 2016
Background

Carvalho et al., 2017
Maunder and Piner 2017

Winker et al., 2018
Kell et al., 2016
• 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
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 residual plot was adapted for integrated age-structured models
• Prediction skill diagnostics (hindcasting) utilizing the Fisheries Library in R (FLR)...

• Recent example of application in IOTC

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

- **Hypothesis (Model Config)**
  - Yes: Run Test on CPUE + Lths
  - No: Check for CorrelatedParms

- **Results Realistic?**
  - Yes: Residual patterns
  - No: Invert Hessian?
    - Yes: Profile critical parms
    - No: Converge Level?

- **Interpret Results:**
  - Either accept hypothesis 1 or create alternative hypothesis
Study Case

North Atlantic shortfin mako shark, and white marlin

Diagnostic tests


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 (>> 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 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.
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.

![Graph showing spawning output over years with shaded confidence intervals and 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 (and versions) used for each diagnostic.

<table>
<thead>
<tr>
<th>Diagnostic</th>
<th>Stock</th>
<th>R(^1) and FLR(^2)</th>
<th>r4ss(^3)</th>
<th>Stock Synthesis(^4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>JABBA-residual plot</td>
<td>Shortfin mako</td>
<td>R (version 3.3.3)</td>
<td>version 1.24.0</td>
<td>(version 3.24U)</td>
</tr>
<tr>
<td>Runs test of residuals</td>
<td>Shortfin mako</td>
<td>R (version 3.3.3)</td>
<td>version 1.24.0</td>
<td>(version 3.24U)</td>
</tr>
<tr>
<td>Retrospective analysis</td>
<td>Shortfin mako</td>
<td>R (version 3.4.4)</td>
<td>version 1.30.2</td>
<td>(version 3.24U)</td>
</tr>
<tr>
<td>R(_0) likelihood component profile</td>
<td>Shortfin mako</td>
<td>R (version 3.4.4)</td>
<td>version 1.30.2</td>
<td>(version 3.24U)</td>
</tr>
<tr>
<td>Jitter</td>
<td>White marlin</td>
<td>R (version 3.4.4)</td>
<td>version 1.35.3</td>
<td>(version 3.30)</td>
</tr>
<tr>
<td>ASFM</td>
<td>Shortfin mako</td>
<td>R (version 3.4.4)</td>
<td>version 1.30.2</td>
<td>(version 3.24U)</td>
</tr>
<tr>
<td>Prediction skill (hindcast)</td>
<td>Shortfin mako</td>
<td>R (version 3.4.4)</td>
<td>version 1.30.2</td>
<td>(version 3.24U)</td>
</tr>
<tr>
<td></td>
<td>FLR</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Stock Synthesis

r4ss

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

- **R₀ likelihood component profile**
  - SS_profile(); SSplotProfile; PinerPlot()

- **Jittering**
  - SS_RunJitter(); Ssgetoutput(); Sssummarize(); SSplotComparisons()
# Retrospective analysis with Stock Synthesis

- Example application
- SWIS ROA FSIC stock assessment
- Stock Synthesis (tested in version 3.10.05 for Windows)
- #R (tested in version 3.10.03)
- #R (tested in version 3.3.3.64 bit)

```r
library(retro)
# install_github("r44s/r44s")
```

## Step 1

- Identify the base directory
- `dimname = "C:\Users\flop\Documents\MyData"`
- Identify the directory where a completed model run is located
- `dirname.completed.model.run = paste0(dirname.base, '/<model name>')`

## Step 2

- Copy model files from the base run
- `dialog.copy(paste0(dirname.completed.model.run, '*', '_<file>.ss', sep=''))`
- `dialog.copy(paste0(dirname.completed.model.run, '*', '<file>.ss', sep=''))`

```r
# starter file changes to speed up model runs
# Run Display Detail
linen <- grep("# run display detail", starter)
```

### Example function

```r
dwtools::install_github("mkagar/kaputils")
library(kaputils)
```
Mahalo!