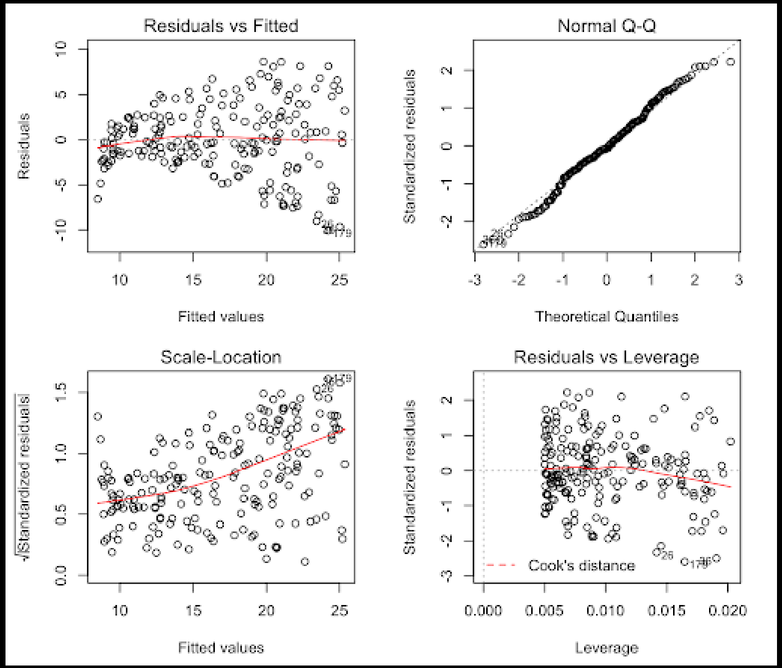
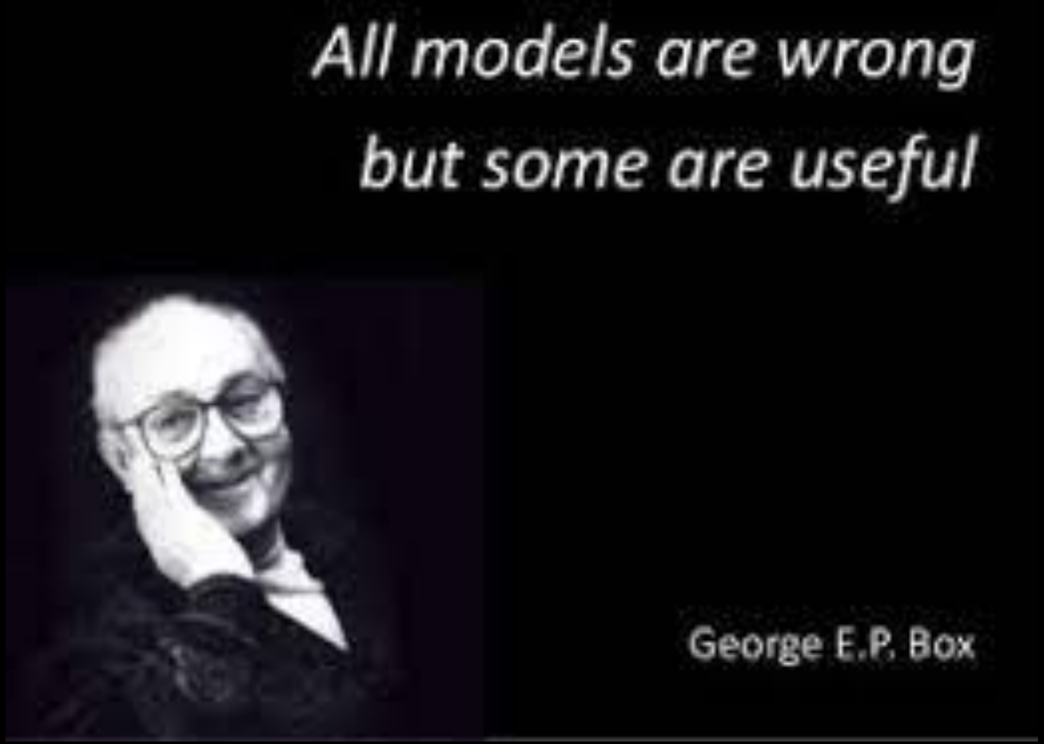
Diagnostics: yesterday, today and tomorrow

- André E. Punt
- School of Aquatic and Fishery Science
- University of Washington

- 3 February 2022

Andre’s opinions: NOT an agreed summary!!





We use diagnostics to minimize “wrong” and maximize “useful”!

- Some history and background
- Data issues
- Model diagnostics
- Get real
- Final thoughts

We don’t have diagnostics (but perhaps should) for “solving the wrong problem”.

SOME HISTORY

It was fifty-odd years ago to day, CEFAS taught the band how to play catch-at-age analysis

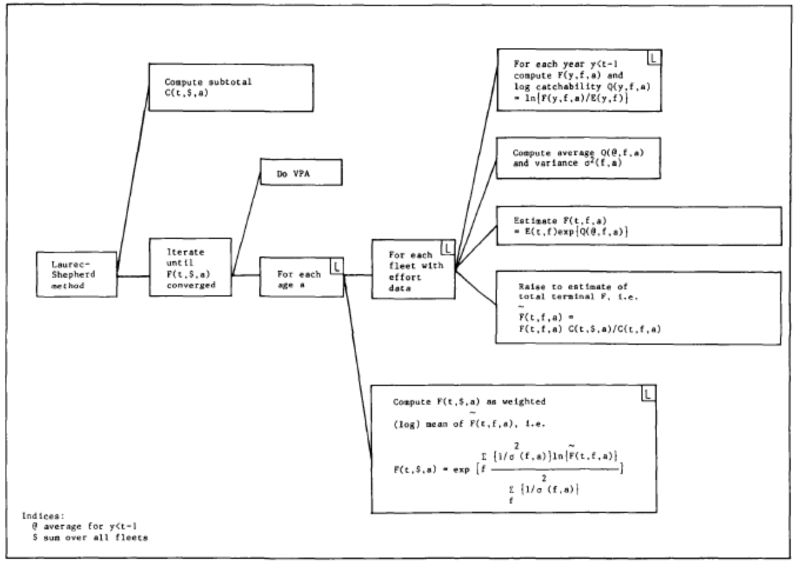


Figure 6. Structure diagram of the Laurec–Shepherd method.

VPAs and production models had little concept of (a) fit diagnostics and (b) model selection. And if they did, it was not formal.

SOME HISTORY

Production model fits weren't all that great!

c) Division 1.6

Model	Minimisation criterion	Year 1981	1982	1983	1984	1985	Relative variation
Babayan et al. (Gulland function-al regression) Equation (6)	"ln cpue"	14,1	13,9	13,9	13,9	14,1	0,007
Lleonart et al. Equation (7)	C	12,5	12,5	12,7	12,8	13,0	0,015
	$\sqrt{C}$	11,6	11,6	11,9	12,0	12,4	0,025
	cpue	9,6	9,1	8,8	8,5	8,3	0,053
	ln cpue	10,9	10,8	11,3	11,5	11,9	0,036
Butterworth-Andrew Equation (10)	C	13,2	13,1	13,0	13,0	13,0	0,006
	$\sqrt{C}$	13,0	12,9	12,9	12,8	12,8	0,006
	cpue	10,6	11,1	11,7	11,9	12,2	0,050
	ln cpue	12,9	12,8	12,7	12,7	12,7	0,006

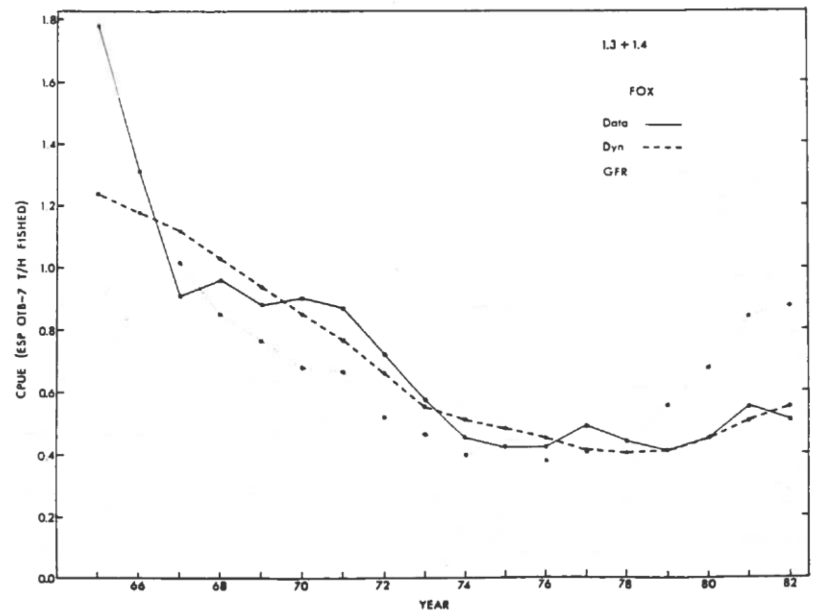


FIG 2. Divisions 1.3 + 1.4, Fox model: plots of the observed cpue time series (solid line) are compared to the expected cpue series for the dynamic model (dashed line) and Gulland functional regression method (dotted line)

Butterworth and Andrew, 1984

But some notion of retrospective / prospective performance evaluation existed.

AND A REGIME SHIFT

Integrated approaches opened the door for:

- More modelling options (what selectivity pattern, growth function, stock-recruitment relationship?)
- Multiple data types

But

- Model mis-specification, model selection issues, contradictory data.

A General Theory for Analyzing Catch at Age Data

DAVID FOURNIER AND CHRIS P. ARCHIBALD

Department of Fisheries and Oceans, Resource Services Branch, Pacific Biological Station, Nanaimo, B.C. V9T 2Y8

FOURNIER, D., AND C. P. ARCHIBALD. 1982. A general theory for analyzing catch at age data. Can. J. Fish. Aquat. Sci. 39: 1195–1207.

We present a general theory for analyzing catch at age data for a fishery. This theory seems to be the first to address itself properly to the stochastic nature of the errors in the observed catch at age data. The model developed is very flexible and accommodates itself easily to the inclusion of extra information such as fishing effort data or information about errors in the aging procedure. An example is given to illustrate the use of the model.

Key words: cohort analysis, virtual population analysis, maximum likelihood estimation, aging errors

FOURNIER, D., AND C. P. ARCHIBALD. 1982. A general theory for analyzing catch at age data. Can. J. Fish. Aquat. Sci. 39: 1195–1207.

L'article qui suit contient une description d'une théorie générale applicable à l'analyse de données sur les prises par âge dans une pêcherie. Pour la première fois, semble-t-il, cette théorie tient compte de la nature stochastique des erreurs que contiennent ces données. Très flexible, le modèle se prête facilement à l'inclusion de données supplémentaires telles que l'effort de pêche ou des renseignements sur les erreurs dans la détermination de l'âge. L'emploi du modèle est illustré à l'aide d'un exemple.

Received December 10, 1980
Accepted April 26, 1982

Reçu le 10 décembre 1980
Accepté le 26 avril 1982

THE PURPOSE OF DIAGNOSTICS

- Help develop a new model or set of models.
- Reject a model (a model must satisfy these to considered for adoption).
- Detect that there is model mis-specification.
- Detect why there is model mis-specification.
- Weight a set of models an ensemble .
- Understand how effective a model will be in providing estimates of management quantities.

Having a cookbook (Terms of Reference) is essential for consistency of review but listing methods is not the same as an automatable system.

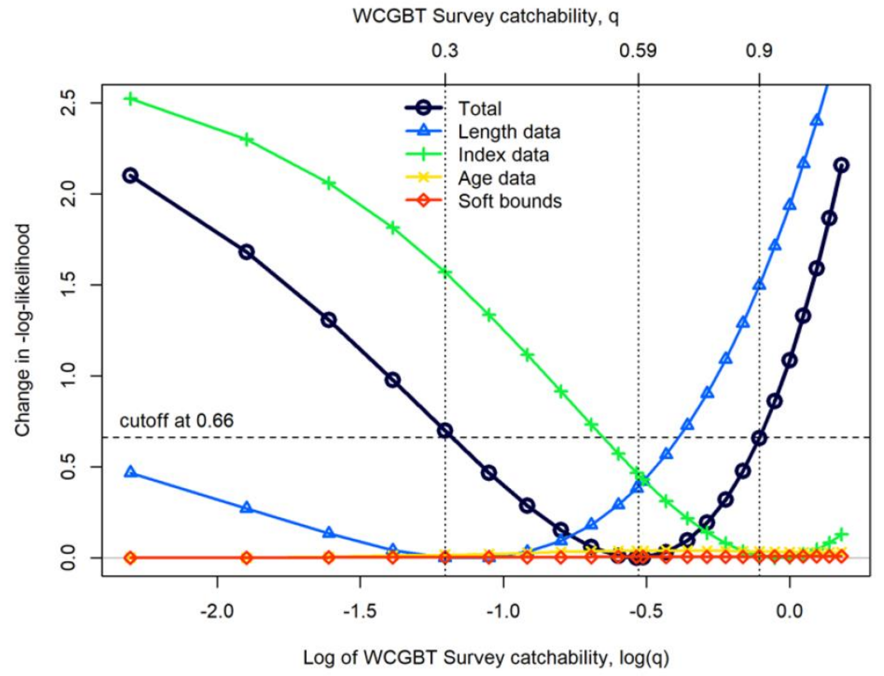


Figure 129. Likelihood profile over log(q) showing contributions of likelihood components. All values are represented as the change relative to the lowest negative log-likelihood for that component within the range of log(q) values shown in the figure.

MODEL DIAGNOSTICS

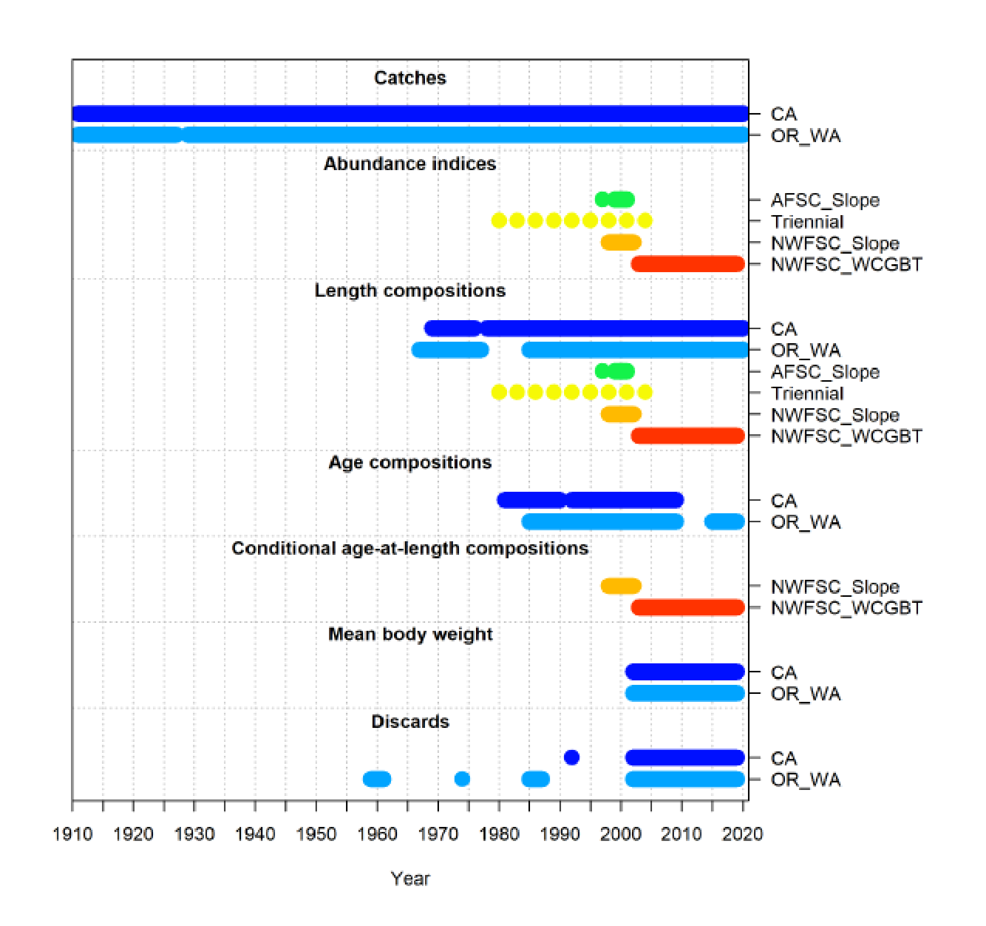
Key principles:

- The “model” is combination of:
  - A model of the population dynamics
  - A model of the observation process
  - How the data are weighted

- The only way we can learn about parameters / make inference is through data (aka the likelihood function). If all you have are priors, you are (at best) inferring the distribution for the model outputs given the priors. Priors cannot / should not be updated without data (or the priors are themselves inconsistent, aka prior checks).

- Diagnostics tell us “something is wrong” but rarely “what is wrong”.

Models are only as good as the data on which they are based



Wetzel and Berger, 2021

DATA DIAGNOSTICS

This is perhaps the most overlooked part of what we do in stock assessment.
Many “data diagnostics” occur prior to any modelling and often involve plots.
Examples of data diagnostics are:

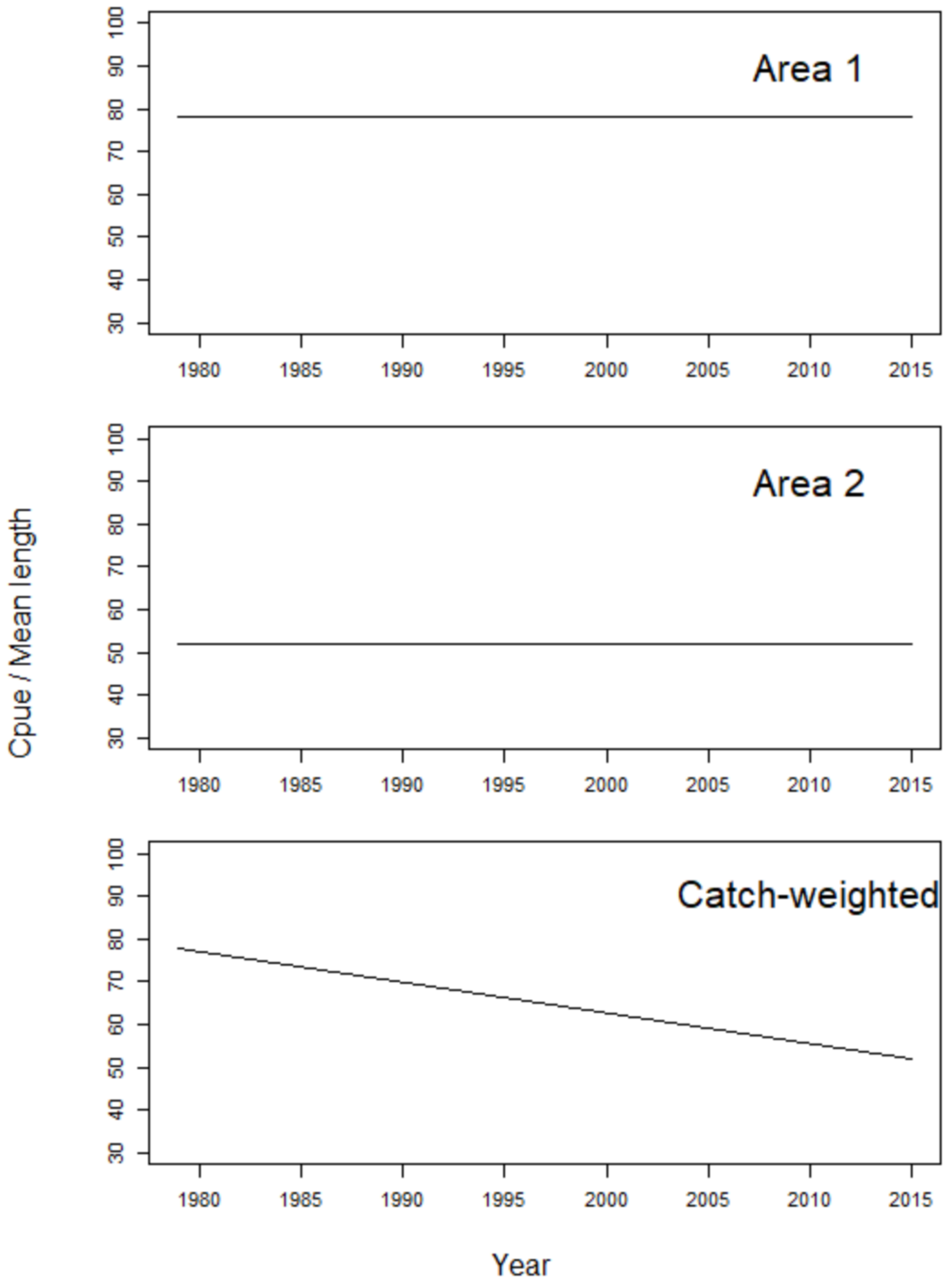
- Looking at the balance of the data (are composition data collected for the areas, seasons, etc to match the catch).
- Are there any outliers / should we implement plus and minus groups?
- How to specify input effective sample sizes (see Thorson’s talk).
- Did we analyze the catch and effort data correctly, e.g.
  - Accounting for the gross lack of independence.
  - Accounting for the fact that fishers “fish where the fish are” (I.e. missing cells are almost certainly not “missing at random”).

Consider the case:

- No change in mean length / CPUE for two areas
- The proportion of the catch in area 1 is declining (e.g. because of high bycatch).
- The net effect is a catch-weighted decline in CPUE and mean length.

What can we conclude:

- The mean length of the catch has declined.
- The abundance of the population has declined.

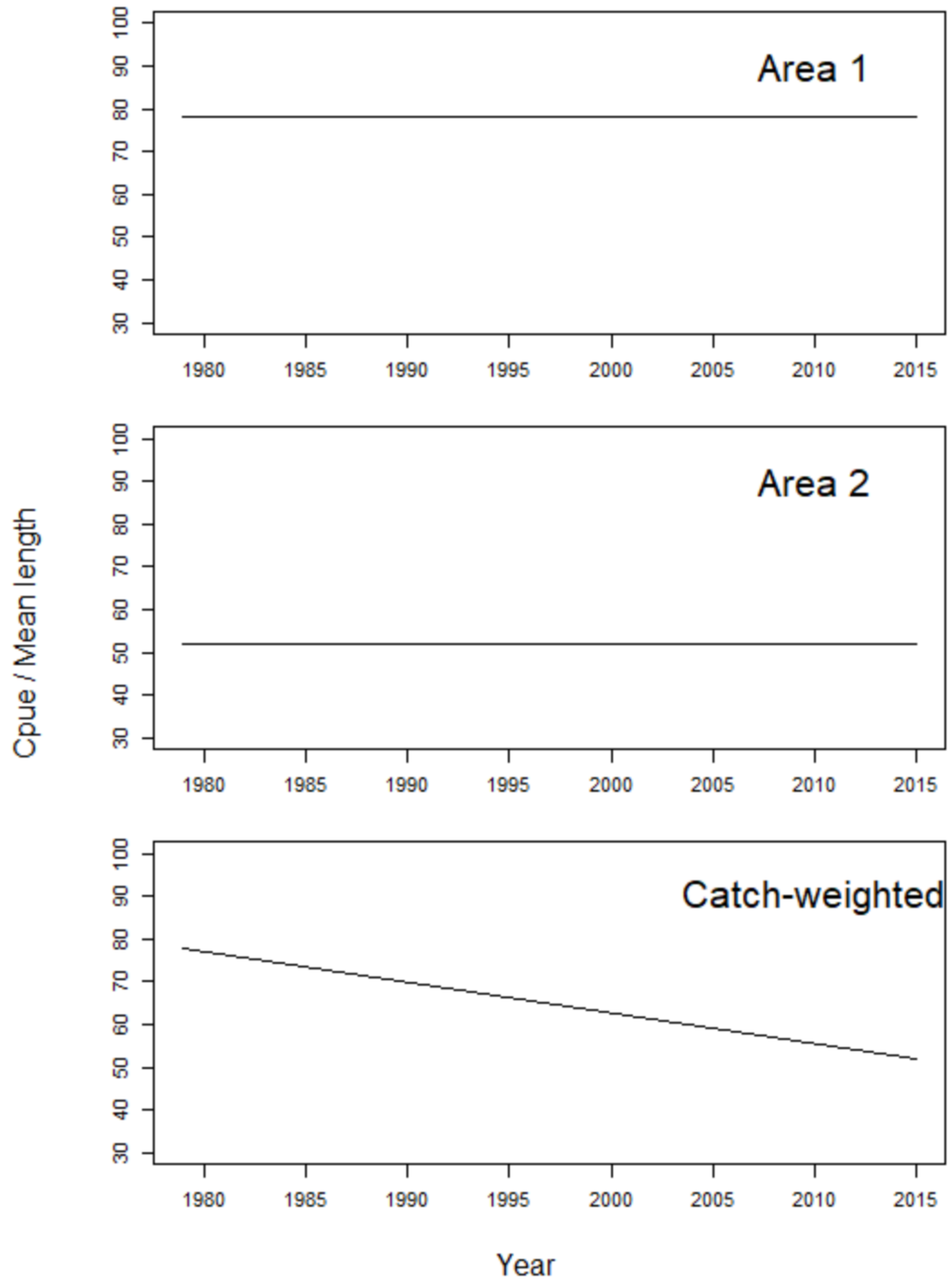


Possible solutions:

- Multiple fleets (one CPUE index per fleet)?
- A CPUE standardization with year*area interactions?
- A spatial model?

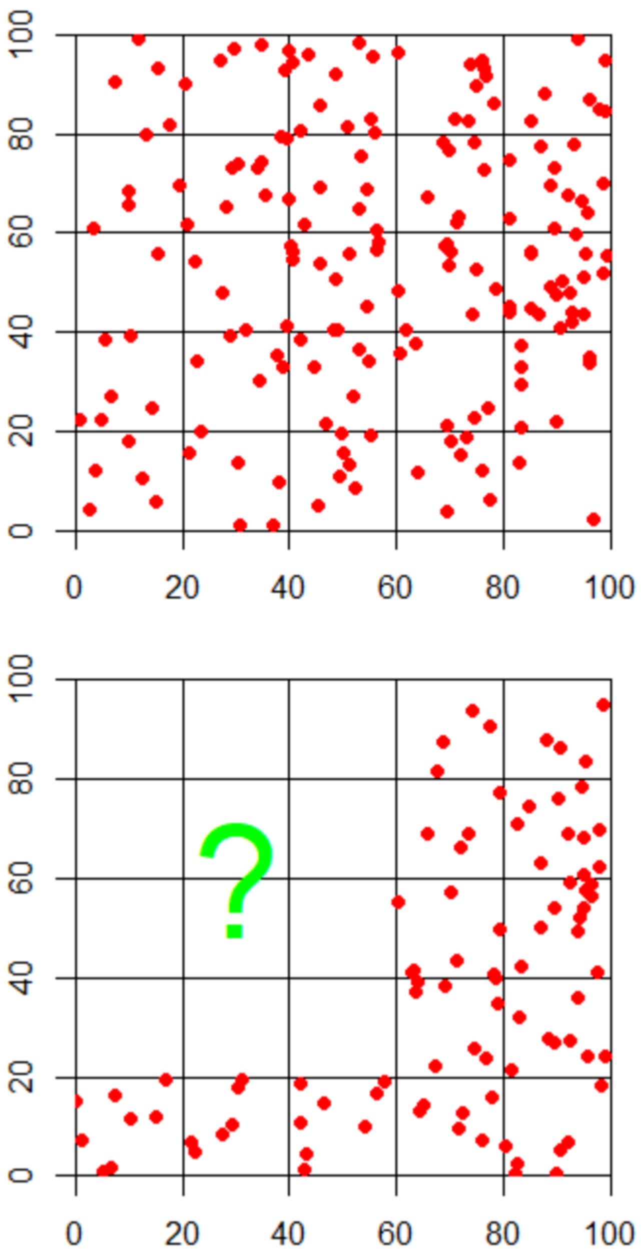
More than a solution:

- Do we have standard diagnostics to understand the nature of our data?
- Lets make sure our focus on model outputs and fit diagnostics is not hiding problems with data?



Changes in spatial distribution:

- The density in the fished areas is the same in two years. What to assume about the area now no longer fished?
  - Assume it is zero: change in abundance is 48%
  - Assume it is the same as the fished area: change in abundance is 0%
  - Assume it is the abundance of highest density grid (e.g. if the unfished area is an MPA)?

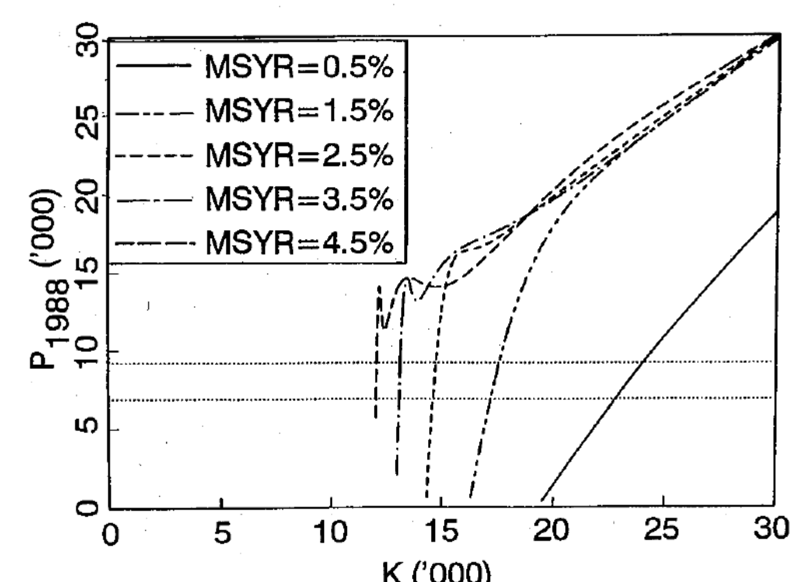


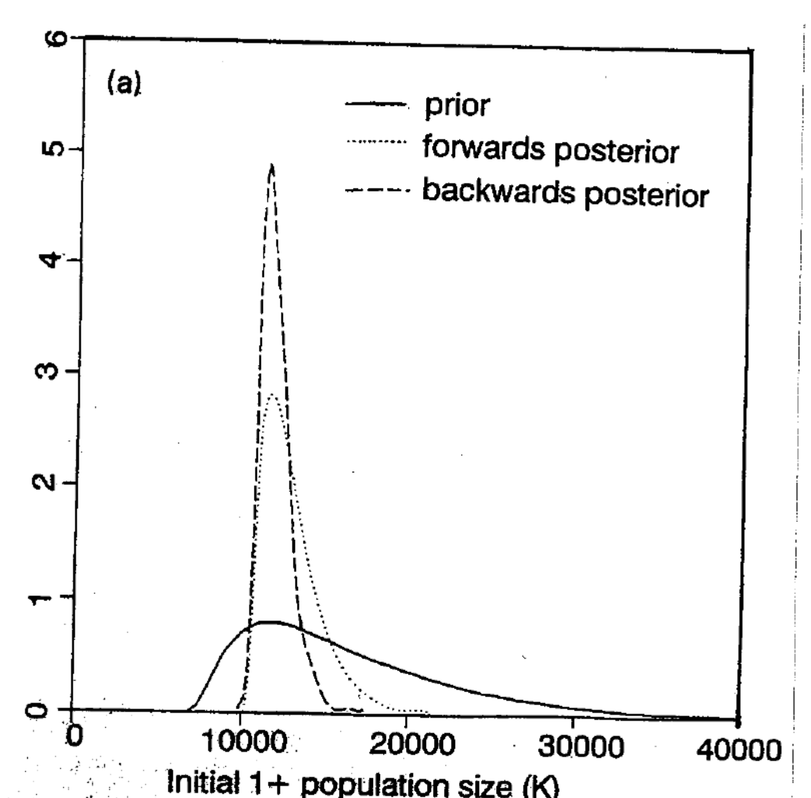
All of our discussions have focused on assessments that involve fitting models to data

What can we say about data-poor (or data-free) methods?

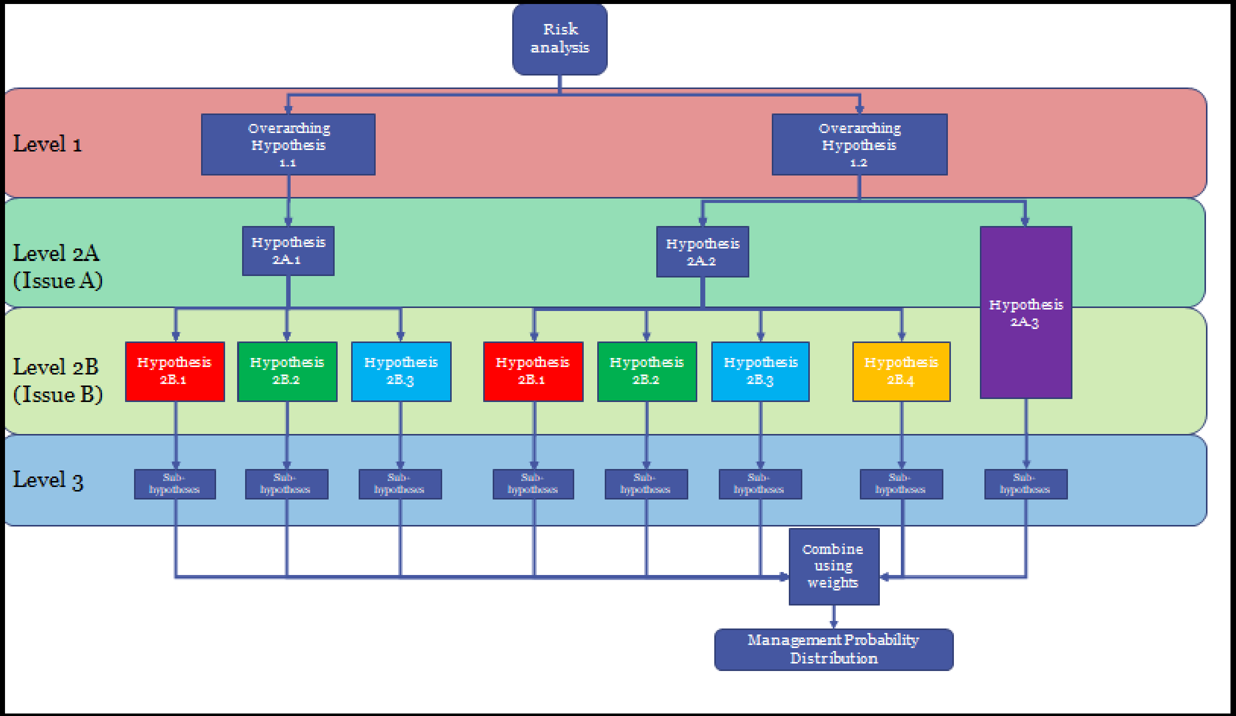
Almost all of the diagnostics are related to models and data – what happens when all we have are priors?

Question: what about prior inconsistencies. Just by postulating a model we can conclude something about carrying capacity….





MULTIPLE MODELS



Once we have multiple models, do we:

- Select one?
- Select many and pool them unweighted?
- Select many and weight them?

Where do diagnostics enter into this?

How to handle:

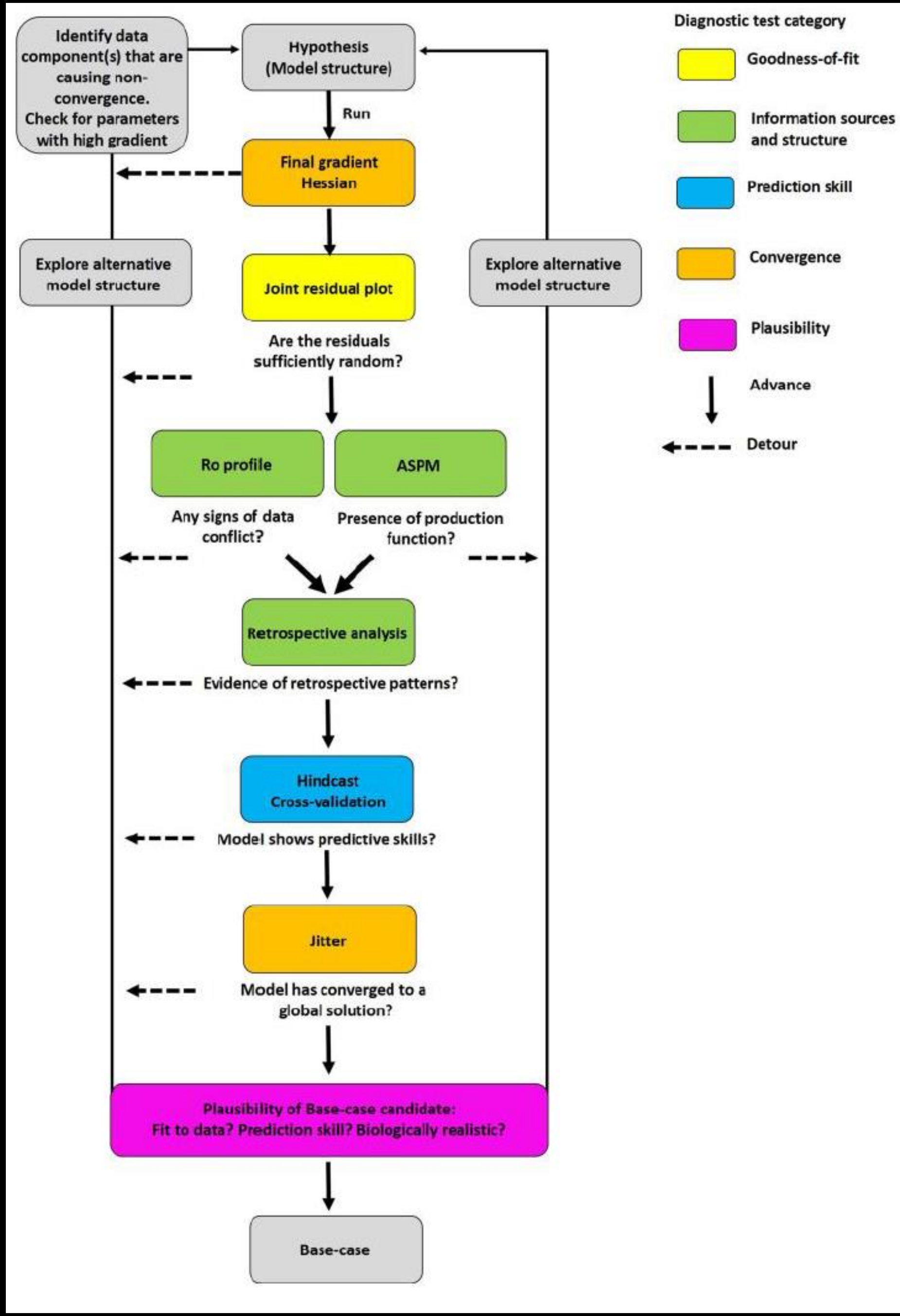
- 4 models?
- 44 models?
- 444 models?

MODEL DIAGNOSTICS

- Convergence diagnostics
- Residual patterns (runs tests, SNDR)
- Effective sample sizes / residual variances
- Retrospective analysis
- Profiles
- ASPM diagnostic
- Catch curve diagnostics
- Hind casting
- Empirical selectivity

Maximum likelihood vs Bayesian

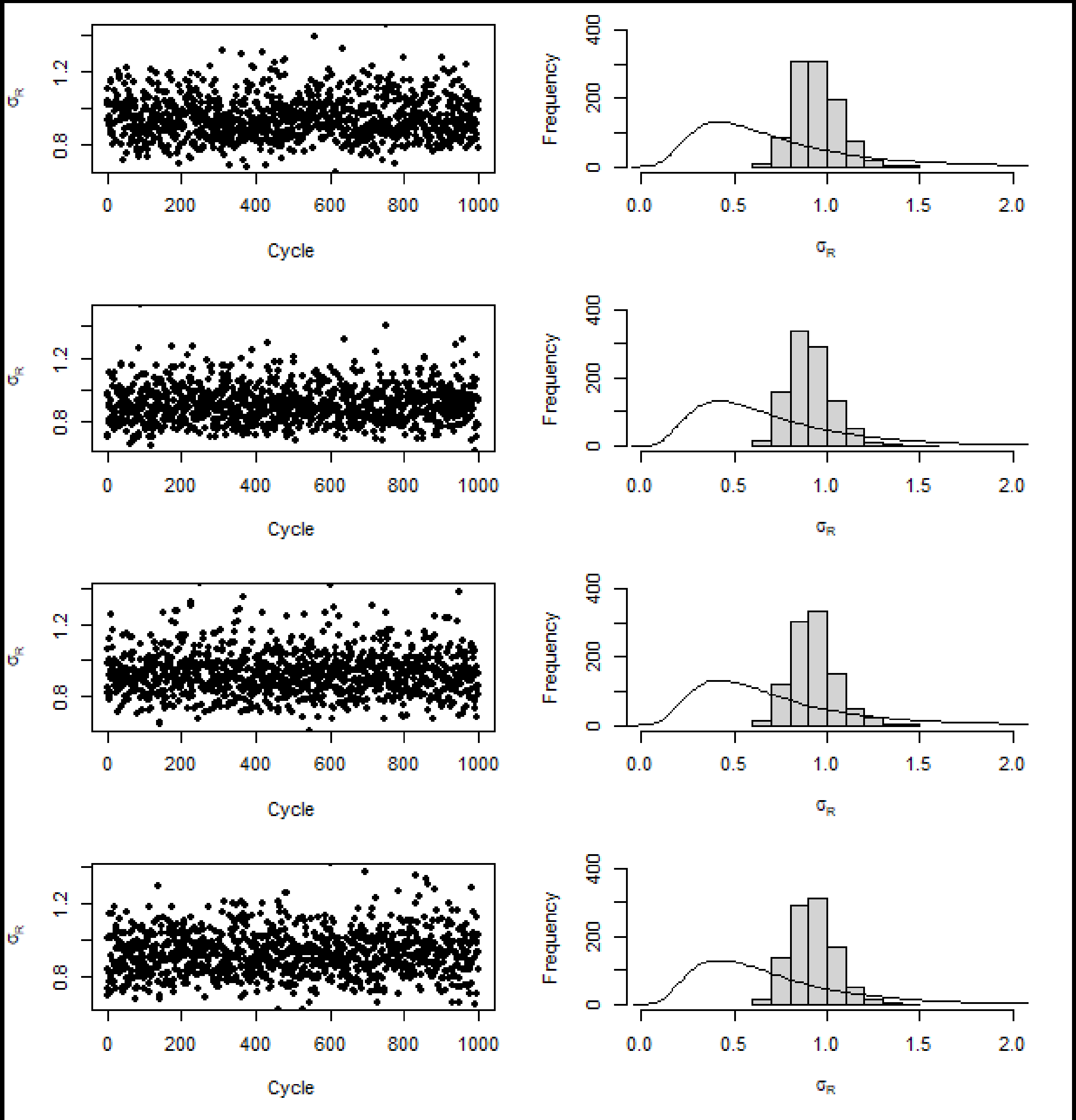
- Prior checks
- Posterior checks



BEST PRACTICES

Convergence (ML or Bayesian) testing

- A non-converged model should ever be allowed to be used for management purposes!
- How many jitters is sufficient to reject a model (if jittering becomes part of model development rather than final model checking)?



Punt et al. Deep Sea Res 2021

BAYESIAN CONVERGENCE CHECKING

What should we see:

- Trace plots (low bar for identifying lack of convergence but the most common "diagnostic" by far)
- Posterior versus prior plots (did the data update the parameters – if no time to do sensitivity analyses).
- MLE versus posterior median comparisons
- Gelman-Rubin statistics based on multiple chains.

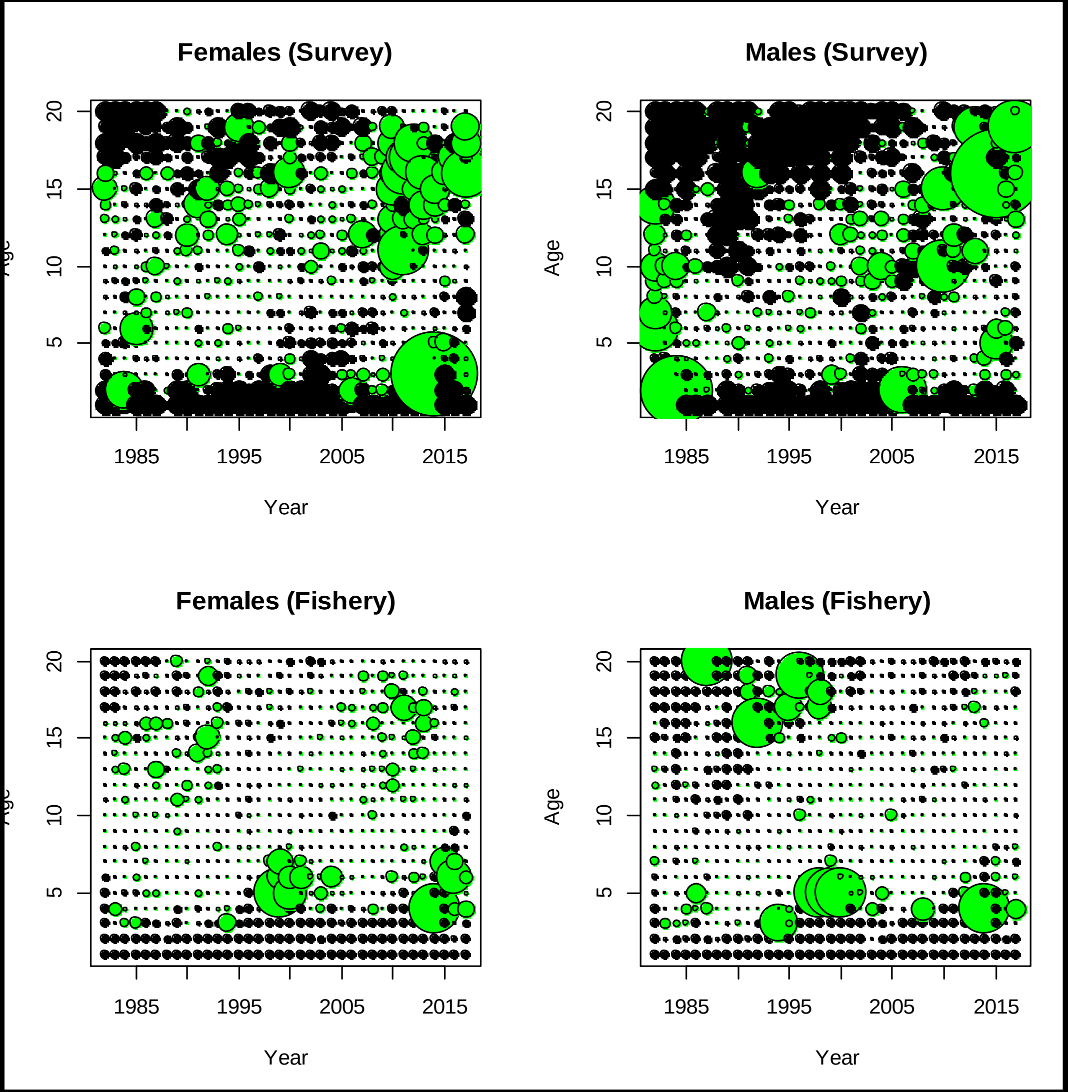
	Development	Reject Models	Detect model misspecification	Detect causes	Weight models	Understand effectiveness
Convergence	Yes	Yes	No	No	No	No

THOUGHTS ON PRESENTATIONS

Residuals:

- Residual patterns (often used during model review but it is never clear “how bad is too bad”). If we keep looking for statistical significant failure we will find it (p-hacking)
- What do you do when nothing seems to improve a residual pattern.

Are these patterns “too bad” (this was published..)

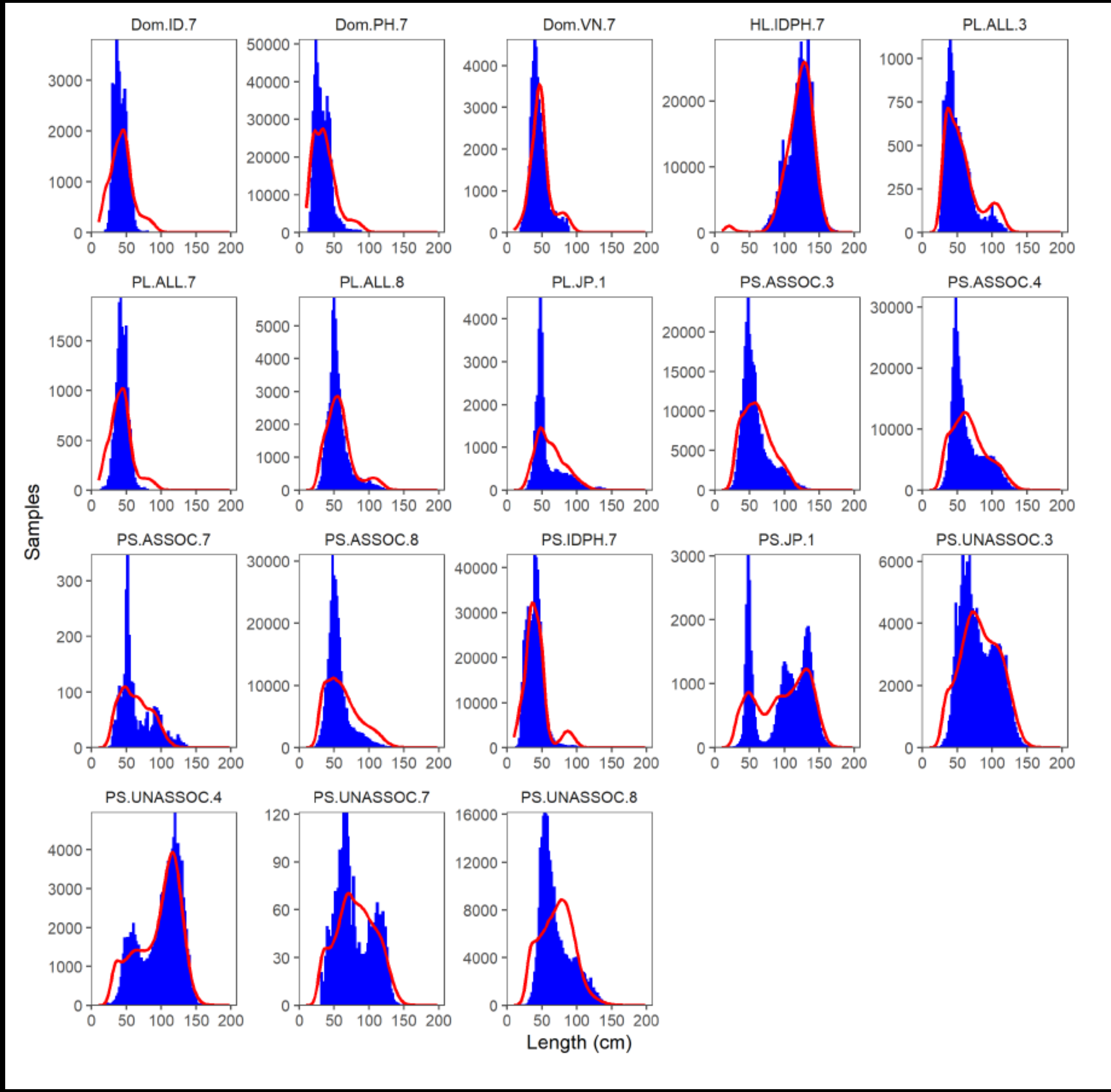


THOUGHTS ON PRESENTATIONS

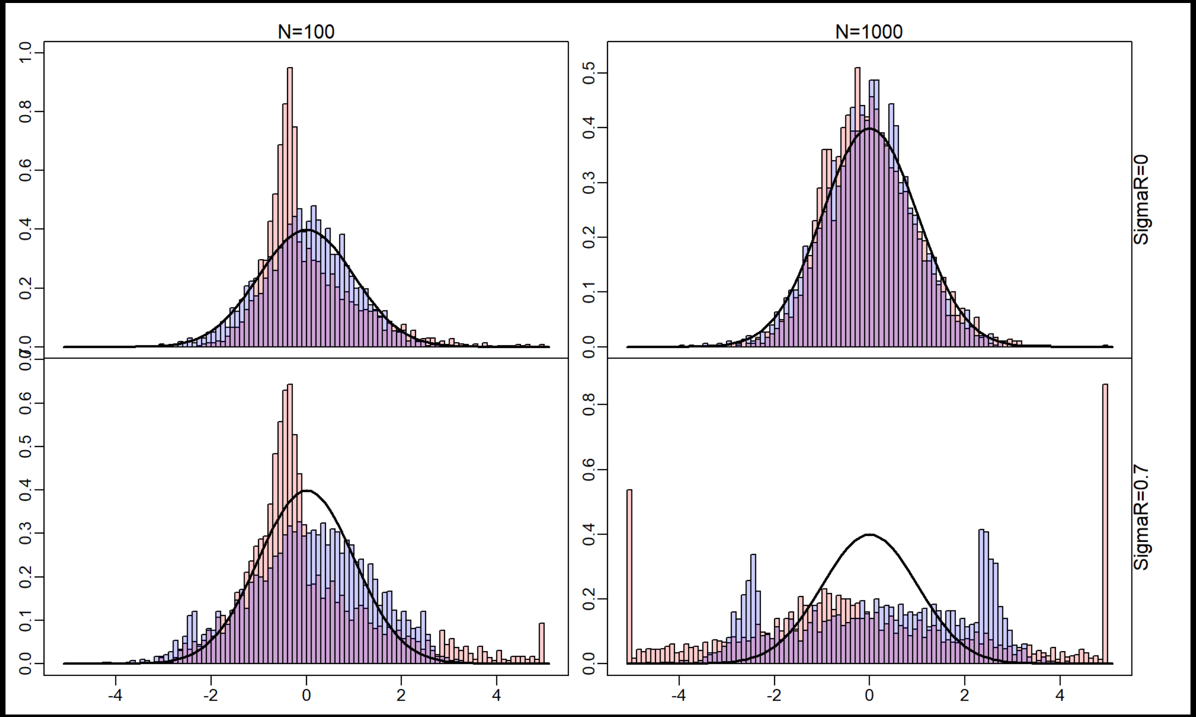
SPC Yellowfin tuna

Residuals:

- Should a failure to fit an aggregated age-/length-frequency be considered terminal (what if the fits to individual years remain poor / wacky).
- Considerable focus is on index patterns and age-/length-compositions. What about tagging data?
- How about indices that are assumed known without error? E.g. environmental data?



THOUGHTS ON PRESENTATIONS



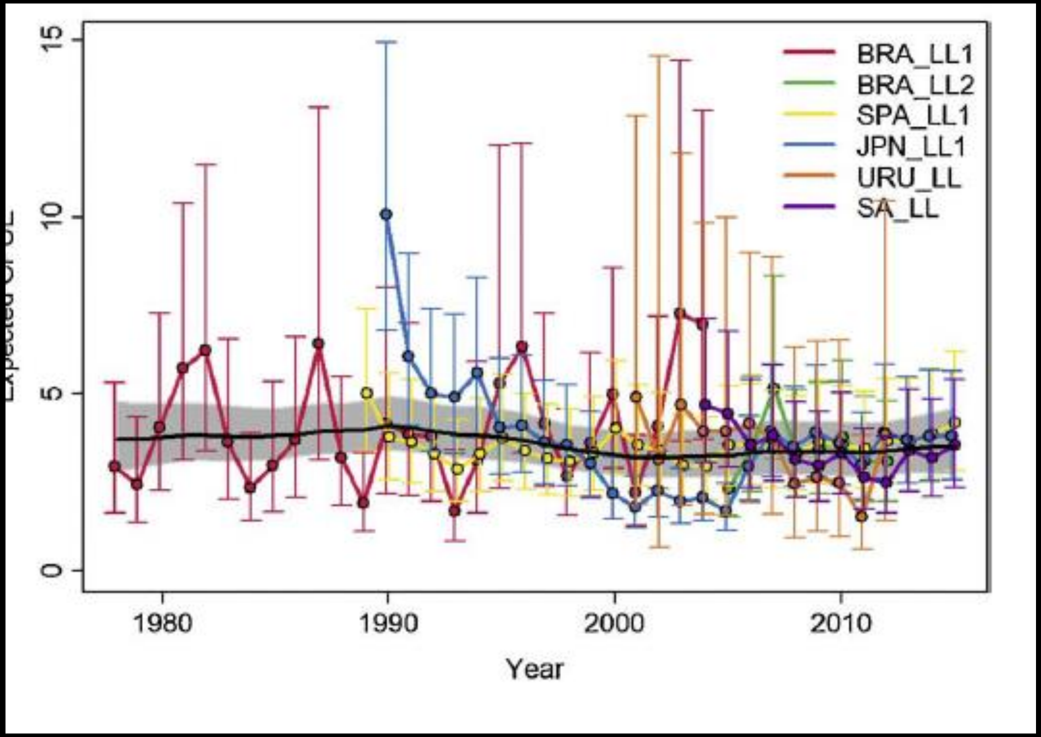
Residuals:

- We should be calculating PIT residuals to include in the standard set (particularly for compositions and tagging data).
- Do Jim Thorson's results imply we may be over-rejecting at present?
- Continue to further explore likelihood-based ways to detect model mis-specification but we also need more simulation studies and validation.

BAYESIAN MODEL CHECKING

Bayesian model checks:

- Very few assessments (based on age- and size-structured models) provide ANY fit diagnostics.
- Low hanging fruit Bayesian-p values for index data.
- Less simple: case-specific diagnostics (c.f. Moran's I for spatial data)



Winker et al. 2018.

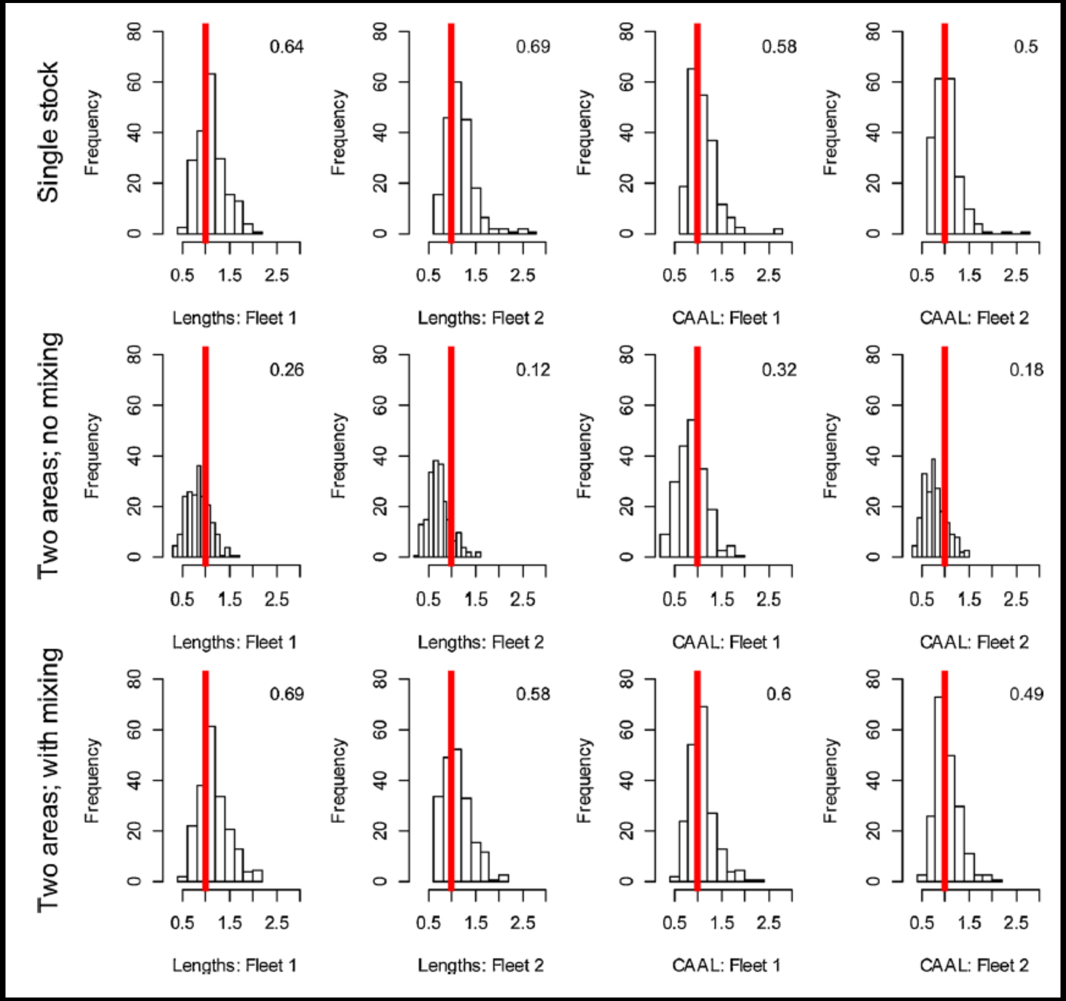
	Development	Reject Models	Detect model misspecification	Detect causes	Weight models	Understand effectiveness
Convergence	Yes	Yes	No	No	No	No
Residuals	Yes	Yes	Yes	Perhaps	AIC?	??

- There is some evidence from simulation studies that residual patterns can detect model mis-specification but identifying causes is not simple.
- Runs tests, etc produce p-values but these do not easily translate into Prob(Model).

THOUGHTS ON PRESENTATIONS

Effective sample size/variances:

- Should we be concerned when the effective sample size is << the input sample size based on the raw data?
- This probably implies that either (a) the input sample sizes have not correctly accounted for the way the data were collected, or (b) the model is too stiff.
- Tuning methods for compositions focus on scaling the input sample sizes but what happens if the relative sizes of the input sample sizes are wrong?



Punt et al., 2021; Fish Ress

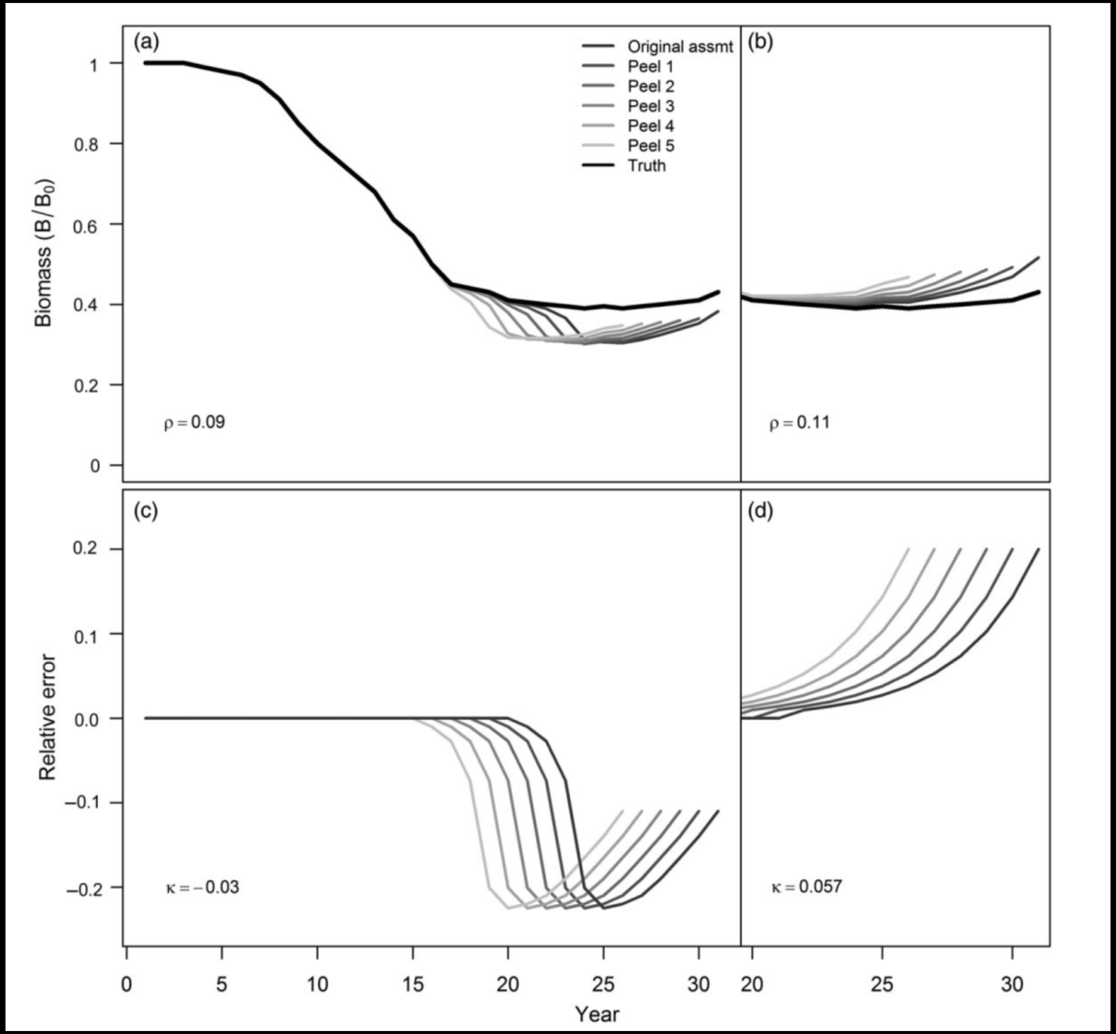
	Development	Reject Models	Detect model misspecification	Detect causes	Weight models	Understand effectiveness
Convergence	Yes	Yes	No	No	No	No
Residuals	Yes	Yes	Yes	Perhaps	AIC?	??
Variances	Yes	Perhaps	No	Yes	??	??

- We know very little about the consequences about input and output variances not matching, but an aim of any model development should be for the input and output variances to match....
- Add process error (properly) does not really add many parameter (if done correctly)

THOUGHTS ON PRESENTATIONS

Retrospective analysis:

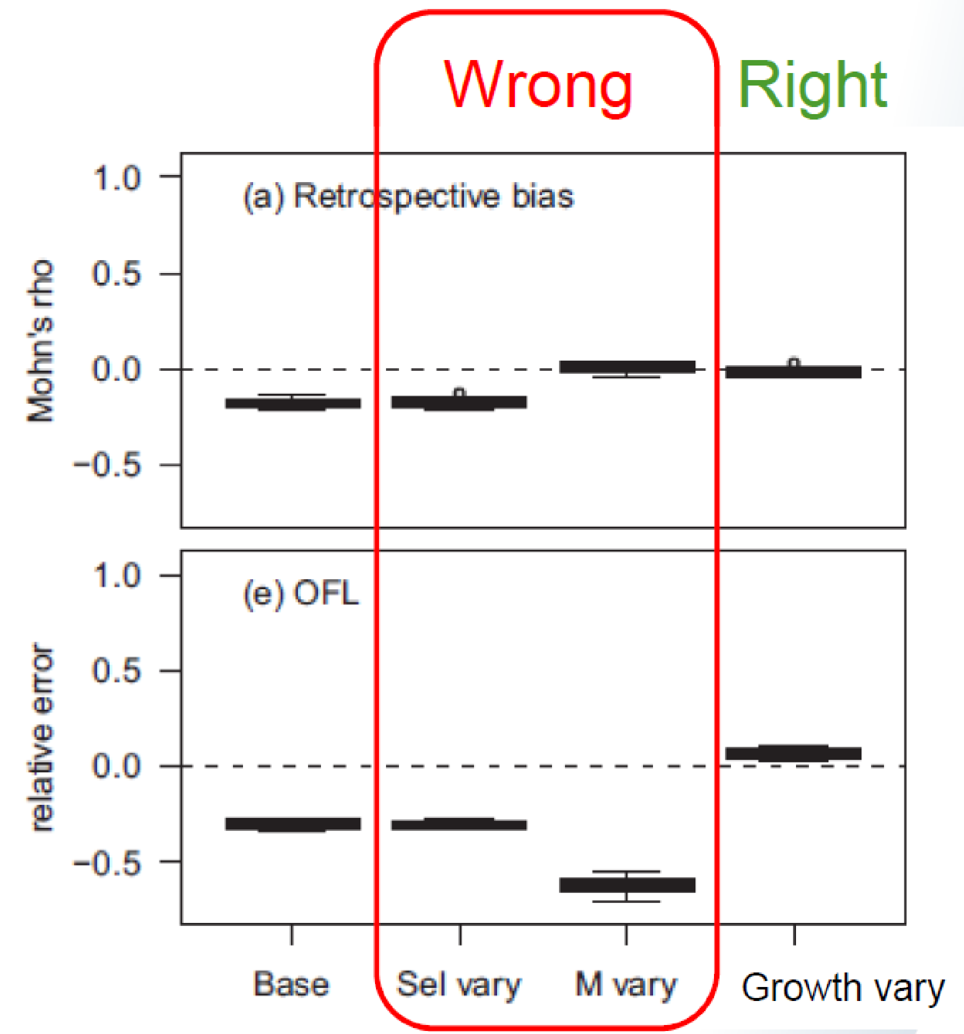
- Should a pattern like this lead to change in the base model / rejection?
  - We need to re-evaluate the Hurtado et al. guidelines to include management consequences
- Yes, there is a problem in some process but why – and what should be done about it?
  - More complex model (random effects)
  - Simpler model
  - Rho-adjustment



THOUGHTS ON PRESENTATIONS

Retrospective analysis:

- Correction by choosing to expand the model will not necessarily correct the model in the right direction.
- Our diagnostics are still very weak in terms of identifying cause of mis-specification.

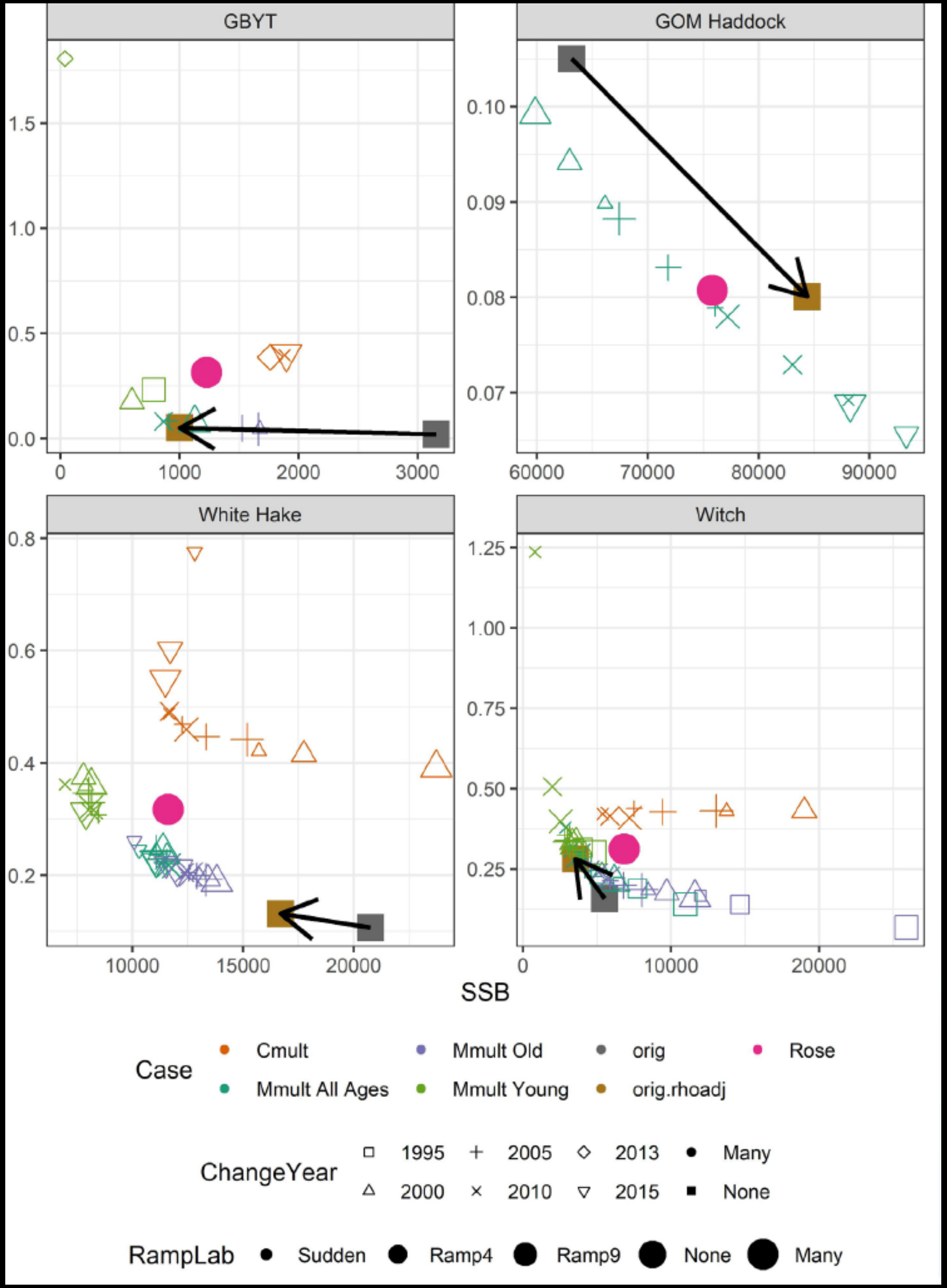


Szuwalski et al. (2018)

THOUGHTS ON PRESENTATIONS

Retrospective analysis:

- The Rose approach:
  - Should be a standard approach but how to automate it?
  - How many models to consider and which?



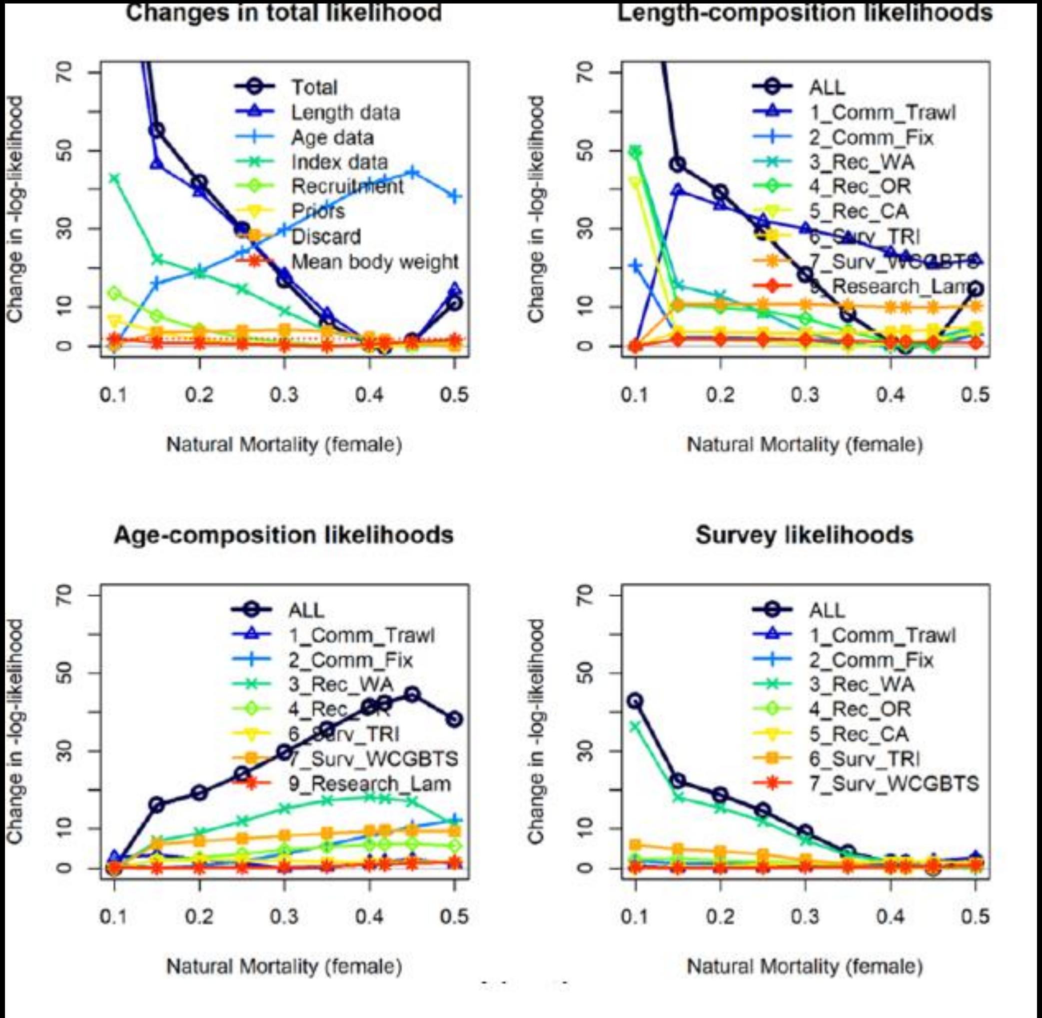
	Development	Reject Models	Detect model misspecification	Detect causes	Weight models	Understand effectiveness
Convergence	Yes	Yes	No	No	No	No
Residuals	Yes	Yes	Yes	Perhaps	AIC?	??
Variances	Yes	Perhaps	No	Yes	??	??
Retrospective patterns	Yes	Yes	Yes	No	Not currently	Yes

- The Rose approach – how to select the models to include in the suite?
- How to translate Mohn's rho into P(model)?
- In principle we know models with strong retrospective patterns probably lead to erroneous management advice (status and catch limits) but how much and does correction fix this?

THOUGHTS ON PRESENTATIONS

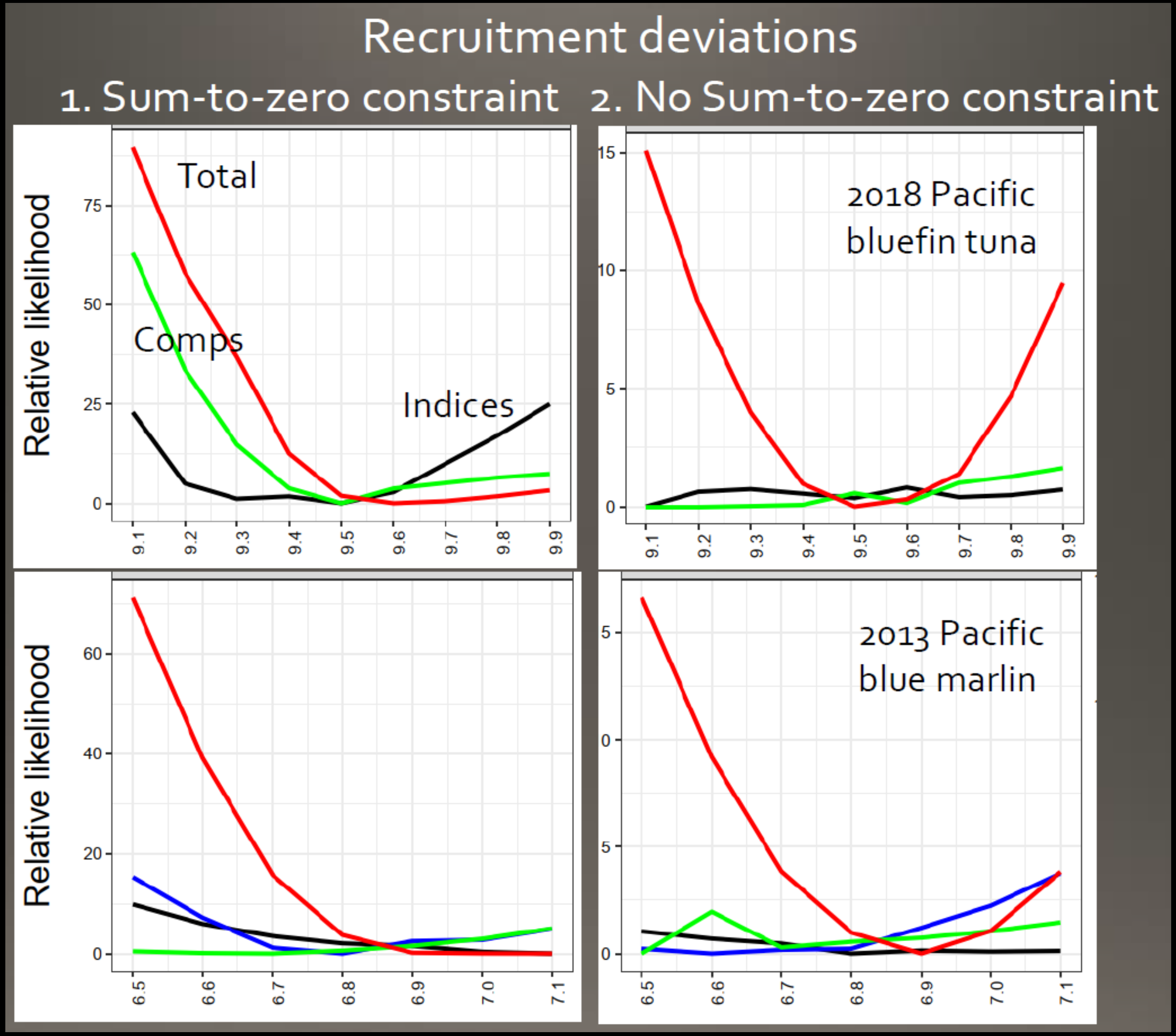
Profiles (including the R_0 profile):

- Generally good to indicate model mis-specification but sometimes the profiles can be confusing (what does a profile that has “conflict” between the index and the recruitment deviations mean?)
- Usually conducted after a candidate base model is selected.
- Should a pattern like this lead to change in the base model / rejection? If not, should a profile ever lead to rejection (or at least alternative models)?



Profiles (including the R_0 profile):

- For R_0 profiles, the no-sum-to-zero component is flatter than when a sum-to-zero constraint is added.
- This may be because the recruitment deviations are being adjusted to mimic a different R_0. This should be checked in future applications.
- Profiles for other parameters should be a standard in assessments to further detect data conflicts (which if the data are correctly collected imply model mis-specification)?

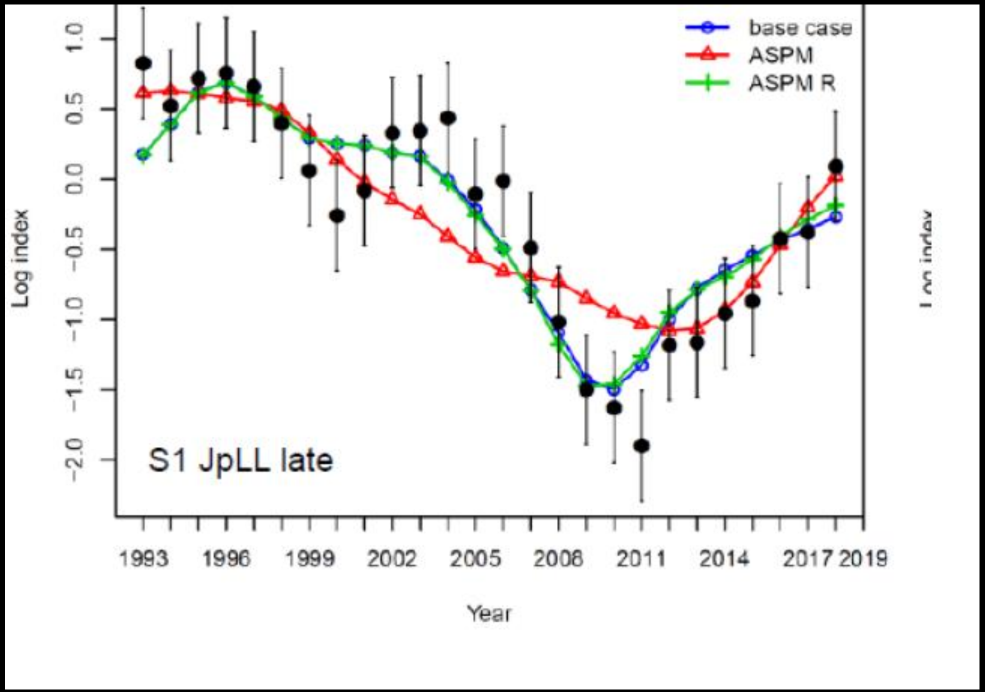


	Development	Reject Models	Detect model misspecification	Detect causes	Weight models	Understand effectiveness
Convergence	Yes	Yes	No	No	No	No
Residuals	Yes	Yes	Yes	Perhaps	AIC?	??
Variances	Yes	Perhaps	No	Yes	??	??
Retrospective patterns	Yes	Yes	Yes	No	Not currently	Yes
R_0 profile	Yes	Perhaps	Yes	No	No	???

THOUGHTS ON PRESENTATIONS

ASPM Diagnostic and catch curve

- The ASPM diagnostic is effective at answering the question of whether the data / model provide evidence for a production function.
- If the ASPM diagnostics does not indicate a production function we need to estimate recruitment deviations.
- A reliable production function should increase forecast skill (but has this been checked using hindcasting)?
- It also indicates the relative information content of the index and composition data.



THOUGHTS ON PRESENTATIONS

ASPM Diagnostic and catch curve

- The catch curve has an enormous type I error rate and appear ineffective as a diagnostic but may help to identify selectivity functions.

Table 7
Percentage of models identified as misspecified by each diagnostic test under different scenarios.

	Self test	Misspecification in selectivity
Diagnostic	CSM(%)	EM_1(%)
SDNR	5	79
Runs test	6	51
ASPM	4	9
Retrospective analysis	0	11
R_o Likelihood component profile	4	5
CCA	91	92

	Development	Reject Models	Detect model misspecification	Detect causes	Weight models	Understand effectiveness
Convergence	Yes	Yes	No	No	No	No
Residuals	Yes	Yes	Yes	Perhaps	AIC?	??
Variances	Yes	Perhaps	No	Yes	??	??
Retrospective patterns	Yes	Yes	Yes	No	Not currently	Yes
R_0 profile	Yes	Perhaps	Yes	No	No	???
ASPM	Yes	No	No	No	No	Yes
Catch curve	Yes	No	No	No	No	Yes

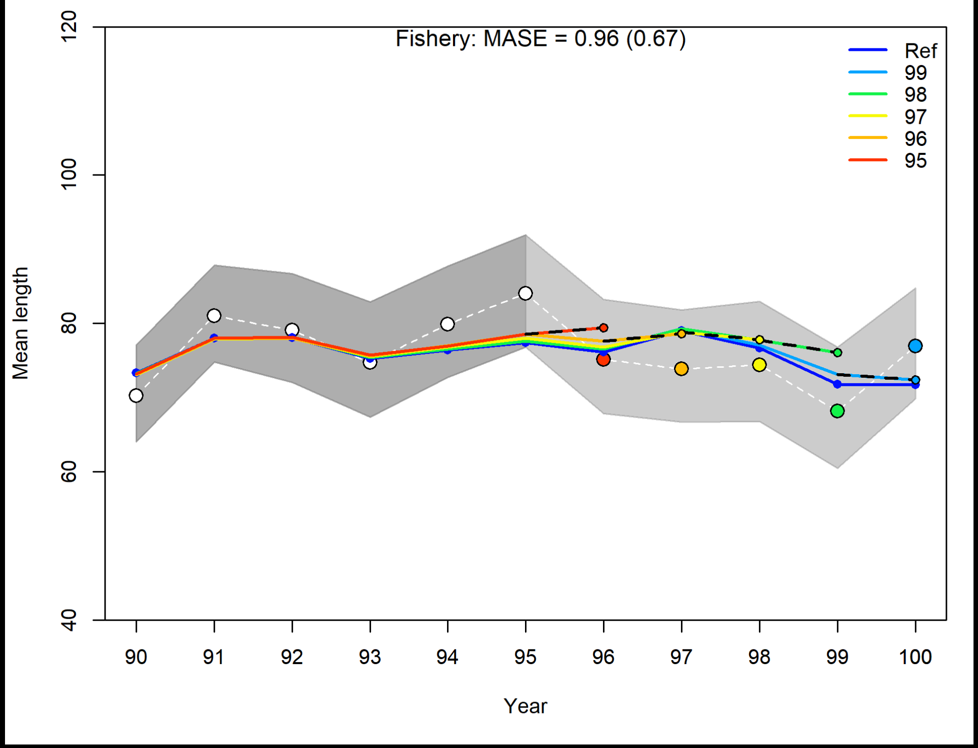
The ASPM diagnostic may be useful during model development and understanding predictive skill (but this needs to be checked). The catch curve diagnostic seems ineffective and should not be used except for model development and understanding purposes.

THOUGHTS ON PRESENTATIONS

Cross validation provides a way to evaluate performance for a model or set of models by dividing the data into a training set and a test set. In principle cross-validation can inform whether there is evidence for overfitting, bias, and whether a model will perform adequately in the future.

Hindcast diagnostic

- This diagnostic involves conducting forecasts of observable quantities (index, metrics of age composition, length composition and tagging).
- There are many ways to apply the method (leave out whole series, leave out data for one fleet, etc) but no best practices.



THOUGHTS ON PRESENTATIONS

Hindcast diagnostic

- Is “no better than an AR-1” an adequate measure of performance?
- We need more guidelines for use of the diagnostic if it is to be used automatically.
- Some simulation testing of the approach given various sources of mis-specification should be conducted. Does the value of MASE change depending on the level of model mis-specification.
- Does MASE < xx denote “no model mis-specification”.

THOUGHTS ON PRESENTATIONS

Hindcast diagnostic

- Can we relate changes in MASEs to specific model mis-specifications?
- What does predicting the following tell us:
  - Index – ??
  - Age/length comps - ??
- Are we concerned that true cross validation is based on leaving data out completely (including during model development)?

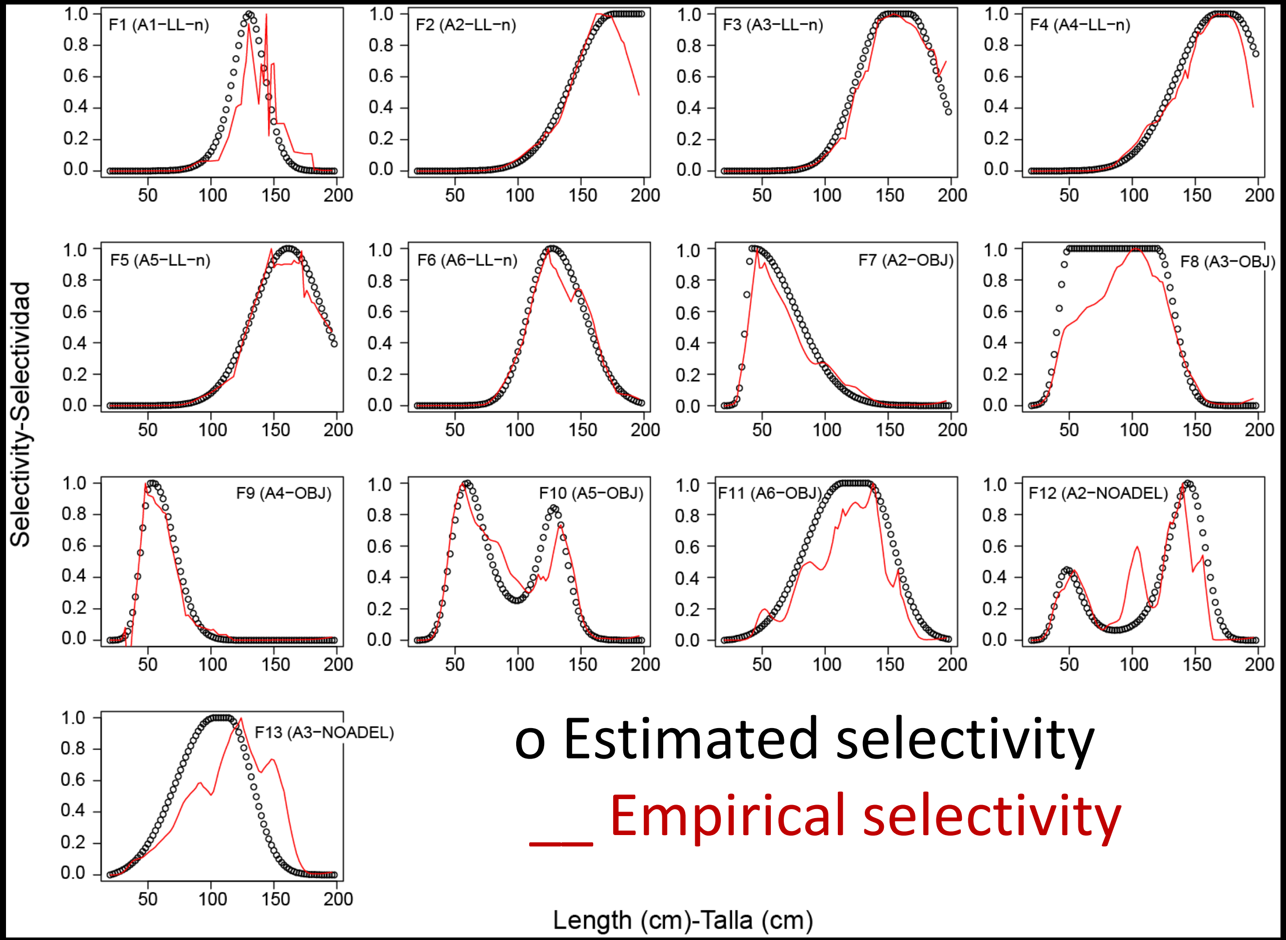
	Development	Reject Models	Detect model misspecification	Detect causes	Weight models	Understand effectiveness
Convergence	Yes	Yes	No	No	No	No
Residuals	Yes	Yes	Yes	Perhaps	AIC?	??
Variances	Yes	Perhaps	No	Yes	??	??
Retrospective patterns	Yes	Yes	Yes	No	Not currently	Yes
R_0 profile	Yes	Perhaps	Yes	No	No	???
ASPM	Yes	No	No	No	No	Yes
Catch curve	Yes	No	No	No	No	Yes
Hindcasting	Yes	No	Perhaps	Perhaps	Not currently*	Yes

This diagnostic can rank models in terms of forecast skill but how do the ranks relate to P(model) ?

THOUGHTS ON PRESENTATIONS

Empirical selectivity tool

- This is not a diagnostic but a way to assist in the construction of models. In principle, it could be a routine part of process of testing final phase.



	Development	Reject Models	Detect model misspecification	Detect causes	Weight models	Understand effectiveness
Convergence	Yes	Yes	No	No	No	No
Residuals	Yes	Yes	Yes	Perhaps	AIC?	??
Variances	Yes	Perhaps	No	Yes	??	??
Retrospective patterns	Yes	Yes	Yes	No	Not currently	Yes
R_0 profile	Yes	Perhaps	Yes	No	No	???
ASPM	Yes	No	No	No	No	Yes
Catch curve	Yes	No	No	No	No	Yes
Hindcasting	Yes	No	Perhaps	Perhaps	Not currently	Yes
Empirical selectivity	Yes	Hopefully not	Perhaps	Perhaps	No	Yes

GE (OR STAY) REAL

The PFMC Guidelines for assessment state: "Evidence of search for balance between model realism and parsimony". What is realism:

- Unlikely parameters ("too high or too low" M, steepness)
- What about population abundance or depletion?
- Selectivity (or movement or ??) patterns that are unrealistic given what is known about the species or fishery.
  - The kill-em vs hide-em debate
- What about cases where the "cryptic" biomass is too large?

The State of Computational Art

	Fully-specified	Automated	Threshold	Notes
Convergence	Yes	Generally	Yes	
Residual patterns	Yes	Yes*	Yes	Move to PIT residuals
Variances	Yes	Yes	No	We really don't what to do fix the problem
Retrospective patterns	Yes	Yes	Yes*	
R_0 profile	Yes	Yes	No	Issues with the recruitment deviations
ASPM	Yes	No	No	Need for recruitment deviations
Catch curve	Yes	No	No	
Hindcasting	Perhaps	No	Yes	Many ways to do this. Also, what does MARE > 1 mean praticticaly
Empirical selectivity	Yes?	Yes	N/A	

THE ULTIMATE DILEMMA & REFLECTIONS

- Should assessments that fail some/most/all diagnostics be rejected?
- It is usually possible to find a model that removes many diagnostic problems (e.g. time-varying zzz) but perhaps by changing the wrong process.
- Should diagnostics such as the ability to predict quantities come into model weighting / what abou
presence of a production function?

In general, allowing selectivity, natural mortality, and growth to vary in the assessment decreased the magnitude of retrospective patterns in estimated spawning biomass, regardless of whether the true time-varying process was allowed to vary. However, the resulting reference points and management advice were sometimes drastically in error when a process other than the true time-varying process was allowed to vary.

Reducing retrospective patterns in stock assessment and impacts on management performance

Cody S. Szuwalski,[1,*] James N. Ianelli,[2] and André E. Punt[3]

[1]*Bren School of Environmental Science and Management, Marine Science Institute, University of California, Santa Barbara, CA 93106, USA*
[2]*National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Alaska Fisheries Science Centre, 7600 Sand point Way NE, Seattle, WA 98115, USA*
[3]*School of Aquatic and Fishery Sciences, University of Washington, 1122 Boat Street, Seattle, WA 98195, USA*

**Corresponding author: tel: 805-893-4058; e-mail: csszuwalski@ucsb.edu.*

Szuwalski, C. S. Ianelli, J. N. and Punt, André E. Reducing retrospective patterns in stock assessment and impacts on management performance. – ICES Journal of Marine Science, 75: 596–609.

Received 6 January 2017; revised 13 July 2017; accepted 16 July 2017; advance access publication 6 September 2017.

Retrospective patterns are consistent directional changes in assessment estimates of biomass in a given year when additional years of data are added to an assessment, and have been identified for a number of exploited marine stocks. Retrospective patterns are sometimes reduced by allowing population processes to vary over time in an assessment, but it is unclear how this practice influences management performance. We simulated stocks in which retrospective patterns were induced by forcing natural mortality, selectivity, or growth to vary over time. We then evaluated the impacts of reducing retrospective patterns by allowing population processes to vary in the assessment. In general, allowing selectivity, natural mortality, and growth to vary in the assessment decreased the magnitude of retrospective patterns in estimated spawning biomass, regardless of whether the true time-varying process was allowed to vary. However, the resulting reference points and management advice were sometimes drastically in error when a process other than the true time-varying process was allowed to vary, and these errors resulted in under-utilizing or over-exploiting the stock. Given the potential for error, identifying the important population processes that vary over time when addressing retrospective patterns should be a priority when providing management advice and may require increased longitudinal life history studies.

Keywords: climate change, management strategy evaluation, retrospective bias, retrospective pattern, stock assessment.

THE ULTIMATE DILEMMA & REFLECTIONS

- We have got very good at using extra variance to account for model mis-specification (as it “resolves” residual variance issues”). Is this smoke and mirrors?

- If the effective sample size says we are only sampling <1 individual perhaps the model is “not entirely correct?
- What does additional variance for a survey mean (perhaps time-varying q / distribution).

FINAL REFLECTIONS

Model diagnostics are (and will remain) a core component of stock assessment science. The key questions are:

- When are diagnostics “too bad” versus “good enough for government work”?
- Do we strive for model with no bad diagnostics, perhaps at the cost of overfitting / bias?
- Can we find diagnostics that assess what is wrong not simply that something is wrong.

When should models definitely be rejected?

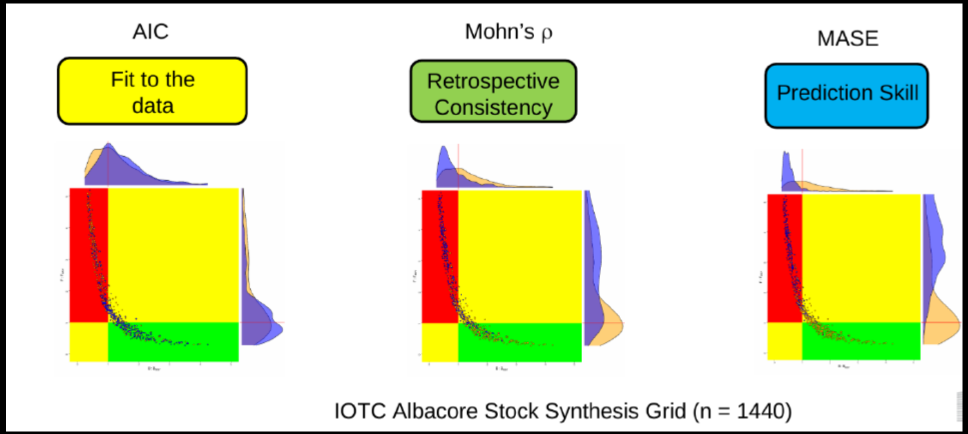
- Lack of evidence for convergence (high gradient; jitter problems; Bayesian convergence failure)
- Biologically implausible results (infinite biomass, steepness =1)
- Clearly (and visually) contradictory data plus R_0 profile?

FINAL REFLECTIONS

The way forward

- Rejected models should be rejected!
- Identify the set of models that “pass” residual analysis (how do we really do for this bubble plots – residual analysis for average age by year, PIT residual by cohort, PIT residual by age).
- Construct a set of models (how to achieve this in balanced way) that reduce retrospective patterns to “near low” Mohn’s rho (a posterior for Mohn’s rho?).
- What is the purpose for sensitivity testing?

FINAL REFLECTIONS



Are we ready to create a weighted ensemble

- IMHO opinion also no.
- We can select models that survive but:
    - AIC weighting is likely questionable (data weights, different data streams?)
    - Mohn's rho and MASE are not measures of relative weight (show me that a weighting scheme for Mohn's rho and MASE based on (say) polynomial functions lead to AIC-type weights and I may have sympathy)

Should prediction skill be used to reject models?

- IMHO opinion probably no. This is a property of the model and should come into understanding model behavior.

Much focus has been on "best assessment" approach – do the same considerations apply to MSE – do we care as much about overparameterization?

NEXT STEPS (KEY)

We have many generic and case-specific diagnostics but:

- We need to conduct a global simulation study to assess type I and power.
- The largest weakness of all diagnostics is the inability to detect which process is mis-specified (detecting there is some mis-specification is easier).
- For rejecting models, we need thresholds for residual patterns, retrospective patterns, etc. (more simulations; including Type I and Type II errors; and cover multiple life histories and data scenarios)
- Hindcast skill is an essential addition to the toolbox, but we need more guidelines for the use of this diagnostic.

NEXT STEPS (OTHER)

$$W(Diagnostics): \frac{\mathbf{W(Diags\ 1) + W(Diags\ 2) + W(Diags\ 3) \dots + W(Diags\ N)}}{\mathbf{Num\ of\ W(Diags)}}$$

- State of the art
  1. Detect convergence problems
  2. Detect mis-specification and data conflicts
  3. Evaluate forecast still
  4. Weight models (I don’t like counting pass/fails) – should we have “bad fail” and account for asymmetric risk.
  5. Identifying mis-specified processes.

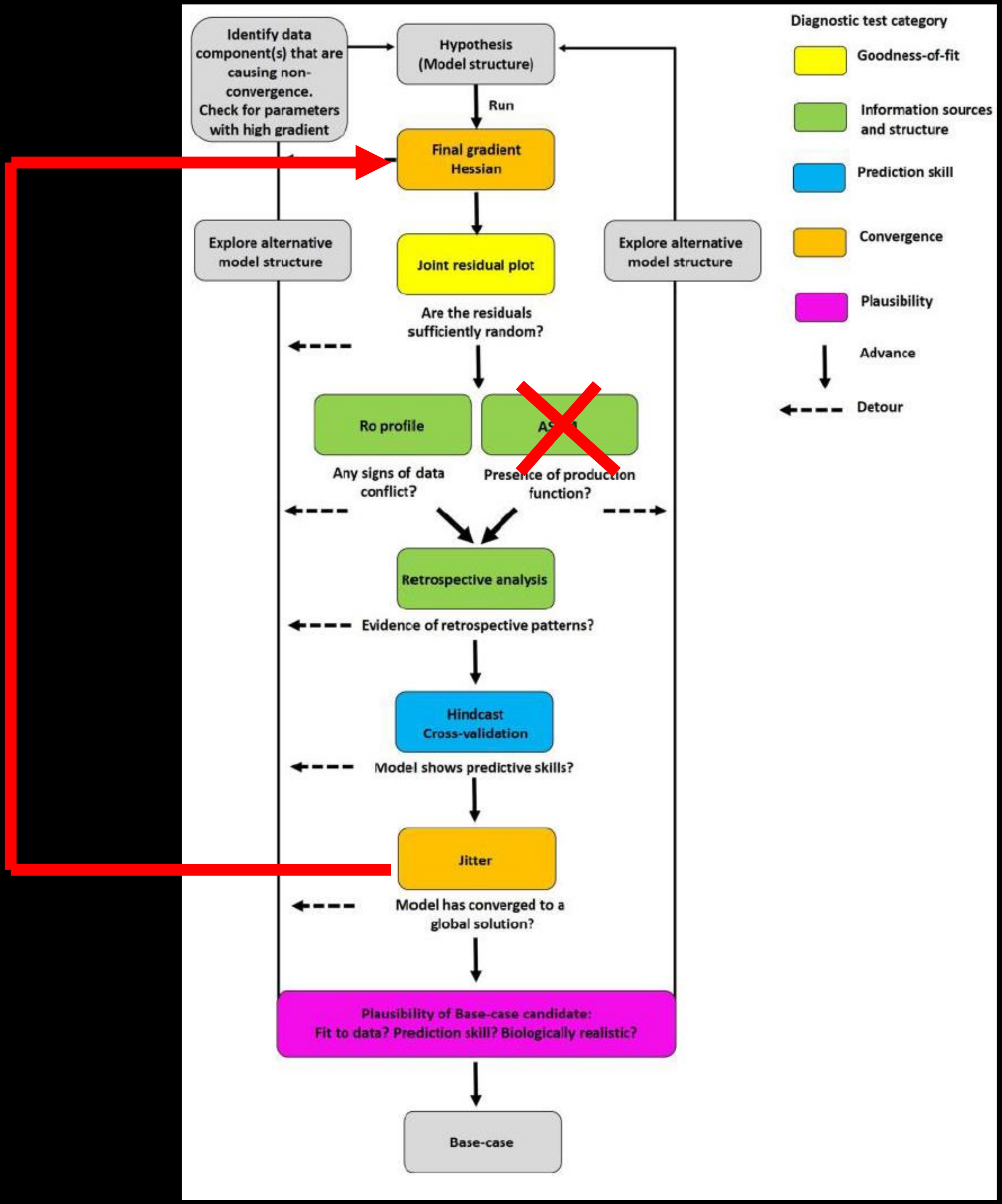
NEXT STEPS (OTHER)

- Develop automatic computation of PIT residuals and further testing to ensure we understand the type I error of our tests?
- Develop age-class, length-class and cohort-specific residual patterns for compositional data.
- Expand understanding of what changes in MASE among models mean.
- Can we a common output format for model output to enhance the chances of common diagnostic tools.
- What is the MASE for a correctly specified model and will a correctly specified always exhibit a production function (I doubt it?)

Andre’s adjusted cookbook

ASPM, MASE & Catch curve -> primarily to understand behavior.

MASE perhaps as model weighting but how?



THANKS

And now to start (continue) disagreeing..