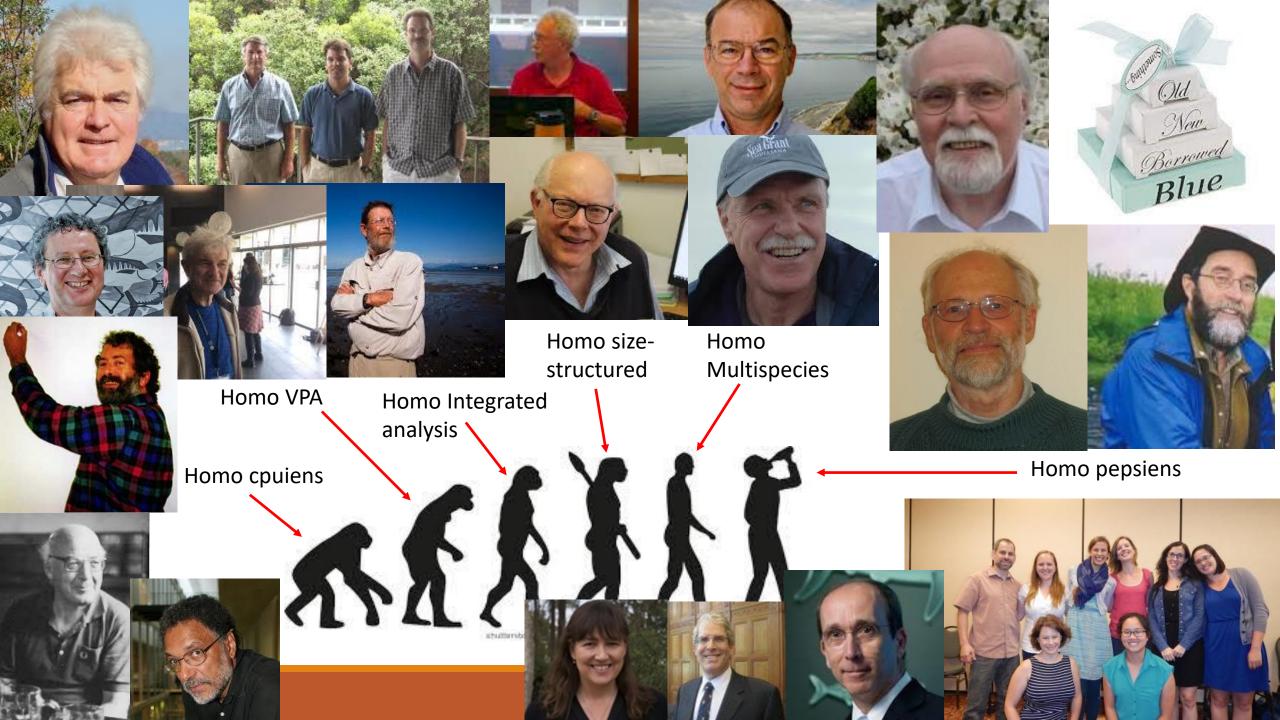


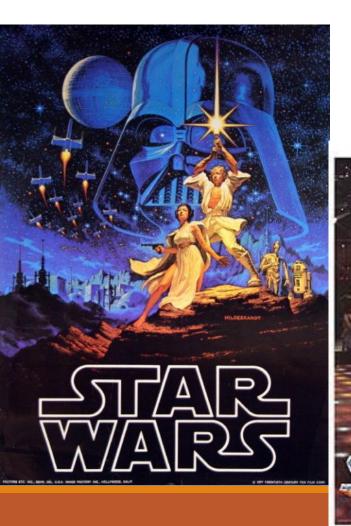


Essential Features of the Next-Gen Integrated Assessment

ANDRÉ E. PUNT UNIVERSITY OF WASHINGTON 2 NOVEMBER 2017



FORTRAN 77





IMPLICIT NONE C*** START OF DECLARATIONS INSERTED BY SPAG INTEGER ACT , LENGTH , NCHAR C*** END OF DECLARATIONS INSERTED BY SPAG INTEGER TODO , DONE , BASE COMMON /EG1 / NCHAR , LENGTH , DONE PARAMETER (BASE=10) 100 IF (TODO.NE.O) THEN ACT = MOD (TODO, BASE) TODO = TODO/BASE IF (ACT.EQ.1 .OR. ACT.EQ.4 .OR. ACT.EQ.7 .OR. ACT.EQ.8 .OR. ACT.EQ.9) THEN CALL BADACT (ACT) GOTO 200 ELSEIF (ACT.EQ.2) THEN CALL COPY LENGTH = LENGTH + NCHAR ELSEIF (ACT.EQ.3) THEN CALL MOVE ELSEIF (ACT.EQ.5) THEN NCHAR = -NCHAR CALL DELETE LENGTH = LENGTH + NCHAR ELSEIF (ACT.EQ.6) THEN CALL PRINT ELSE GOTO 100 ENDIF DONE = DONE + 1CALL RESYNC GOTO 100

ENDIF 200 RETURN END

SUBROUTINE OBACT (TODO)



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

For me, this was the most significant paper I read as a graduate student. It introduced me to the next 30-odd years of my life:

- Methods that use multiple data types
- Introduction of "state space" formulations for models in fisheries.
 - Separation and modelling of both process and observation error.

Today, most general approaches to stock assessment are state-space formulations (or approximations thereof).

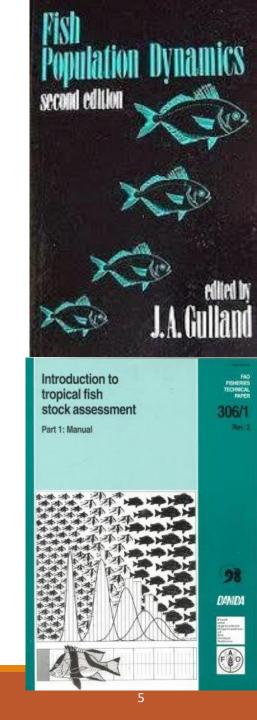
What is "state-of-the-art" assessment (today)

Structure

- Single species and stock.
- Age- or size-structured.
- Multiple fisheries or surveys (modelled as areas-as-fleets).
- Fitted to multiple data sources (usually index, length-/agecomposition, growth data, discard).

Technical stuff

- "Random effects" treated as "penalized parameters".
- Coded in ADMB (or similar).
- Some measure of uncertainty (usually asymptotic methods).
- Data weights "tuned" using *ad hoc*-ish methods.



Reminder areas-as-fleets:

We model $N_{y,a}^s$ and assume that difference spatially in age- / size-structure is due to selectivity, but fundamentally the population is fully mixed.

A full spatial model will consider $N_{y,a}^s$ and assume that difference spatially in age- / size-structure is due to **population structure** and selectivity. The population is still fully-mixed but locally.



The N-marix

Important caveats

Most assessments don't require a lot of features

- High feature assessments are those with complex data sets.
- The complexity of an assessment often relates to the data available rather than the biology (don't make an assessment more complex than it needs to be).
- There is a trade off between bias and variance but also be realism and specification error.

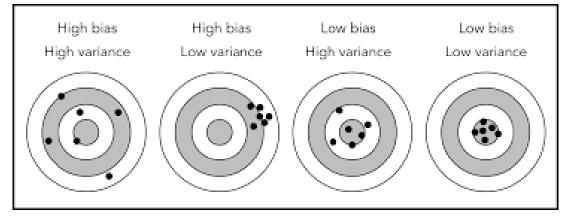
Some features pertain to how assessments will be used

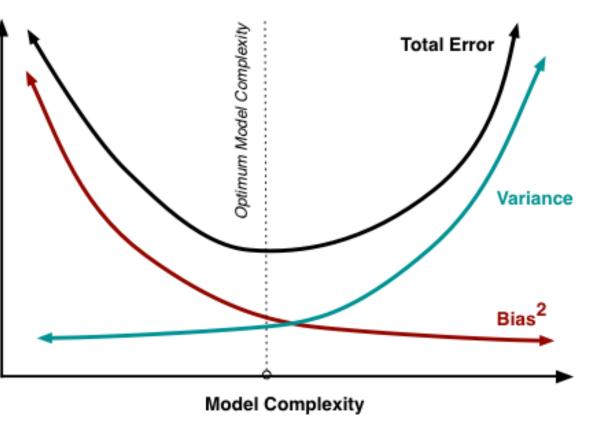
 Case-specific control rules can add substantially specification complexity (e.g. the SS forecast file) but may be essential (why does the US want OFL, ABC and ACL, but it does..)



The originator of 14th century AICc

- Each assessment will have a different "sweet spot" that minimizes total error.
- We need:
 - Modelling frameworks that allow us to create models given data for a particular case.
 - Methods to work out where the "sweet spot" is.





Pinterest.com

www.snoek.ddns.net

6 challenges on 16 slides

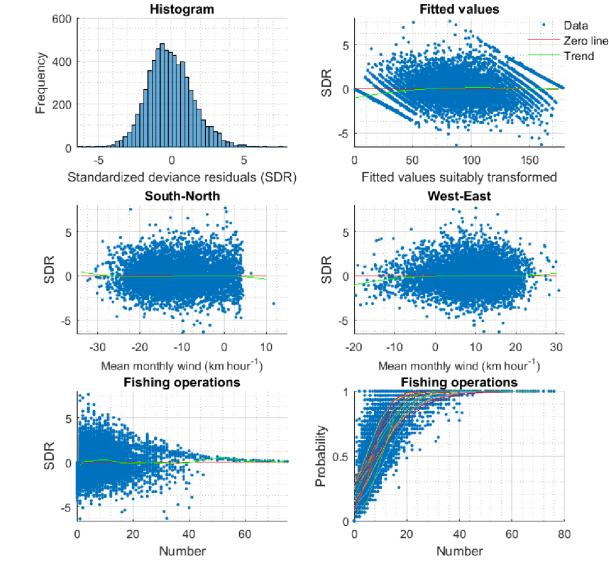




The six challenges for features

- Multiple "stocks" and species in one assessment.
- Spanning the range from data-rich to data poor situations.
- Age-length models
- Mark (Maunder)-recapture data
- Dealing with random effects and data weighting "correctly"
- Simulation and Management Strategy Evaluation



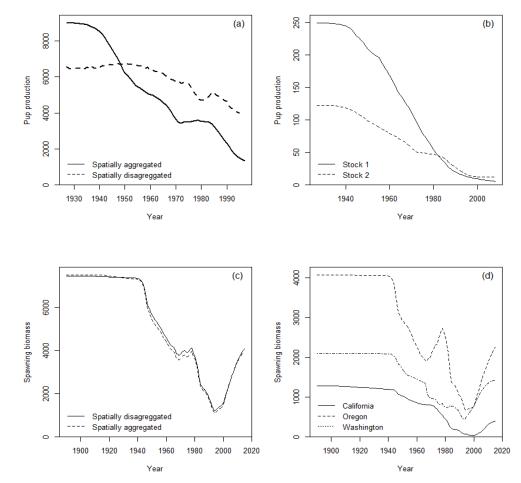


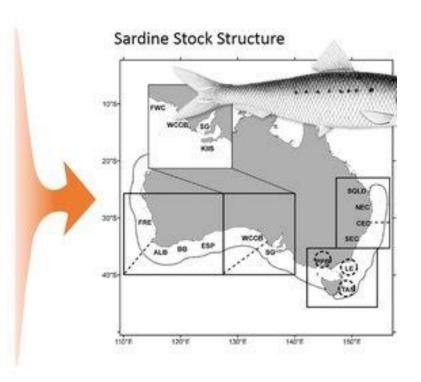
Most assessment packages (including SS) assume that there is a single stock (i.e. density-dependence depends on a single spawning biomass) [SS allows for "growth patterns" (which can be area-assigned)].

CASAL allows for multiple populations, with some parameters shared among modelled stocks.

When there are multiple stocks, it is desirable for data-rich stocks to inform data-poor stocks:

- parameters should be "shared" / "mirrored" among stocks; and
- values for parameters should be allowed to be subject to a common prior among stocks.



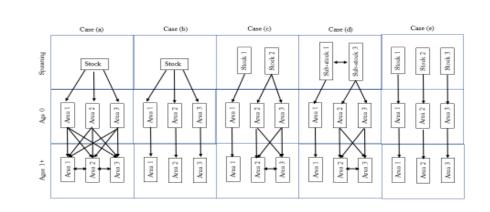


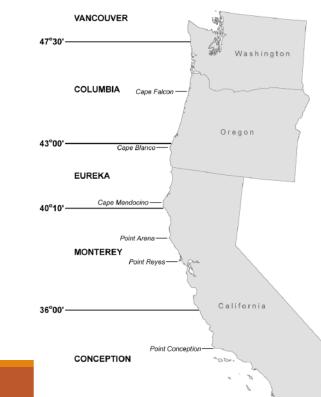
Desired stock/species structure features

- Multiple "stocks" and species.
- Area- or stock-specific density-dependence.
- Data may relate to multiple stocks.
- Parameters can be mirrored between stocks and fleets.
- Density-dependence in distribution and movement.
- Sex-specific distribution of recruits.
- Ability to place a penalty / priors on trends in F among stocks and species.
- Ability to add data on stock mixing.

Challenges for spatial structure

- Multiple populations that mix (and between which there is dispersal).
- Spatial (rather than population-level) density-dependence.
- Clines in density / biological parameters across what constitutes one "biological" population (a current major challenge).
- Calculation of reference points (e.g. F_{MSY}) in presence of multiple stocks that mix and when biological parameters vary spatially.
- Modelling of movement is still limited in most packages (including those that allow for spatial structure).



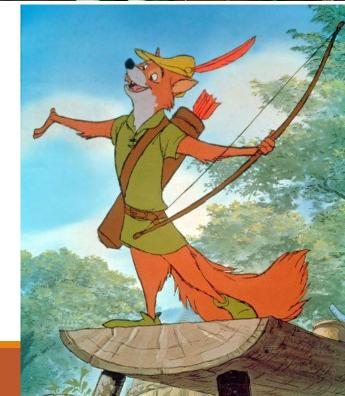


Scaling from data-rich to data-poor

- Data-rich assessment methods:
 - Multiple data types and many parameters and processes
 - Fitted: maximum likelihood and Bayesian approaches
- Data-poor assessment methods:
 - Production models, catch-only methods
 - Catch-curve analysis
 - Fitted: Bayesian approaches; *ad hoc* fitting
- Data-free assessment methods:
 - Catch-only methods
 - Yield-per-recruit and spawning biomass-per-recruit

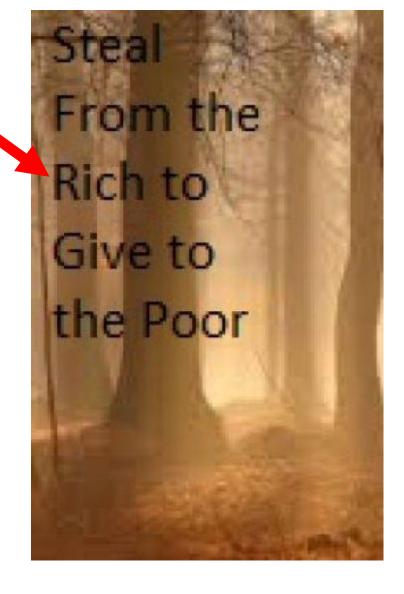
I am focusing here on methods for estimating F and B (in absolute terms and relative to reference points); control rules are a different story.





Desirable features

- Data-rich -> data-poor in the same framework
 - For example, Stock Synthesis, -> XSSS -> SSS
 - Most likely Bayesian to allow propagation of uncertainty (avoid fixing too many parameters; data poor methods should have higher estimates measures of uncertainty)
 - More use of (data-based informative) priors on key parameters (we talk but seldom act)
 - Input files should "scale" with the complexity of the model.
 - Able to inform data-poor assessments from data-rich assessments (aka the Robin Hood approach).



Wattpad

Data

Evaluating the role of data quality when sharing information in hierarchical multistock assessment models, with an application to Dover sole

Samuel D.N. Johnson and Sean P. Cox

Abstract: An emerging approach to data-limited fisheries stock assessment uses hierarchical multistock assessment models to group stocks together, sharing information from data-rich to data-poor stocks. In this paper, we simulate data-rich and data-poor fishery and survey data scenarios for a complex of Dover sole (*Microstomus pacifiats*) stocks. Simulated data for individual stocks were used to compare estimation performance for single-stock and hierarchical multistock versions of a Schaefer production model. The single-stock and best-performing multistock models were then used in stock assessments for the real Dover sole data. Multistock models often had lower estimation errors than single-stock models when assessment data had low statistical power. Relative errors for productivity and relative biomass parameters were lower for multistock assessment model configurations. In addition, multistock models that estimated hierarchical priors for survey catchability performed the best under data-poor scenarios. We conclude that hierarchical multistock assessment models are useful for data-limited stocks and could provide a more flexible alternative to data pooling and catch-only methods; however, these models are subject to nonlinear side effects of parameter shrinkage. Therefore, we recommend testing hierarchical multistock models in closed-loop simulations before application to real fishery management systems.

Age-length models

- Most stock assessment models are either age-based on size-structured. However:
 - Size-structured models cannot fit to age-composition data
 - Age-structured models can fit approximately to size-composition data (e.g. using morphs / growth patterns in SS)
- Age-length models do exist
 - They address age- and length-selectivity more accurately
 - Inclusion of tagging data in an age-length model would be more natural given most tagged fish are sized but not aged
 Exploring model structure uncertainty using a general stock
 - But they can be quite slow
- Are simpler approaches may be adequate:
 - Platoons
 - Age-slice methods.

Exploring model structure uncertainty using a general stock assessment framework: The case of Pacific cod in the Eastern Bering Sea

Caitlin I. Allen Akselrud*, André E. Punt, Lee Cronin-Fine School of Aquatic and Pishery Sciences, University of Washington, Seattle, WA 981 95-5620, USA

ARTICLE INFO

ABSTRACT

Article history: Received 11 January 2017 Received in revised form 21 March 2017 Accepted 22 March 2017 Handled by George A. Rose

Keywards: Age-structured model Age-size-structured model Assessment Pacific cod Platoon-structure Size-structured model An assessment framework is developed that allows analysts to conduct stock assessments for fish and invertebrate stocks based on age-, size- and age-size-structured population dynamics models. The sizestructured model is nested within the age-size-structured model. The framework can use catch, discard, index of abundance, size- and age-composition, conditional age-at-length, mean length-at-age, and tagging data to estimate model parameters. It is used to explore the sensitivity of key model outputs for Pacific cod in the Eastern Bering Sea by applying model configurations that use the same data, same likelibood functions, and same data weighting schemes. Base model configurations using the three model types all fit the available data adequately, but the age-structured model fits the data better than the size-structured model. Variation in estimates of spawning biomass and the overfishing level was higher among model-types than within model-types. This result highlights the need for assessment analysts to focus more on applying and presenting results for multiple models.

© 2017 Elsevier B.V. All rights reserved.

Perspective: Let's simplify stock assessment by replacing tuning algorithms with statistics

Reherics Resource: Assessment and Monitoring Division, Northwest Fisherics Science: Center, National Marine Robertes Service, NOAA, Seattle, WA, USA



Three "deadly sins" of stockassessment

- Tuning of σ_{R}
- Tuning of data weighting
- Failure to include random effects where they are needed.

A.E. Pant	Stock assessments are important to sustainable ocean management, but developing assessments remains time-
	consuming despite increased computation power and access to shared software. Improved efficiency in devel-
	oping stock assessments could allow an increased rate of new assessments, increased attention to biological
i model	mechanisms in existing assessments, or accelerated testing of existing methods. I argue that the efficiency of the
anent	stock-assessment enterprise is hindered by a reliance upon ad hoc "uning algorithms" that are conducted in-
ion	dependently of standard parameter estimation. I present three examples where tuning algorithms are widely
£	used: (1) determining the variance of recruitment, (2) bias-correcting recruitment deviations, and (3) de-
	termining the effective sample size for compositional data, and summarize why each tuning algorithm was
	originally developed. I then review recent research showing that each task can be replaced with parameter
	estimation involving random effects. Finally, I explain how model development, peer review, and model testing
	would each be improved if tuning algorithms were replaced by parameter estimation, and outline the steps
	required to transition existing stock assessments to modern parameter estimation involving mixed effects.

Stock assessment methods based on **TMB** or **Bayesian methods** are not subject to these concerns to quite the same extent as our current penalized likelihood methods. Lets fix this **now**.

James T. Thorson

ARTI

Handled by Kerwards:

Mixed-effect Hierarchical Stock assess

Rommitmon

Statistical Science 2016, Vol. 31, No. 2, 259–274 DOI: 10.1214/16-STS552 In the Public Domain

Close-Kin Mark-Recapture

Mark V. Bravington, Hans J. Skaug and Eric C. Anderson

Term	Definition	Status
n	Sample size (# captured individuals)	Observed
P	Capture probability at given sampling occasion	
	Quantities related to a captured individual	
у	Time (year) of birth	Latent
x	Place of birth or capture	Latent/Observed
t	Time of capture	Observed
z	Covariate vector at time of capture	Observed
a	Age	Latent/Observed
L	Length	Observed
	Population dynamics and demography	
Ν	Total population size (males and females)	Parameter
φ	Individual survival probability	Parameter
α	Age of maturity	Parameter
β	Per capita average birth rate, $\beta = \mathbb{E}(R)$	Parameter
R	Reproductive output (# offspring) of an individual in a given year	Latent
R ₊	Total reproductive output from all individuals in a given year	Latent
	General	
θ	Vector of all model parameters	Parameter
K	Kinship category (see Table 1)	Latent/Observed

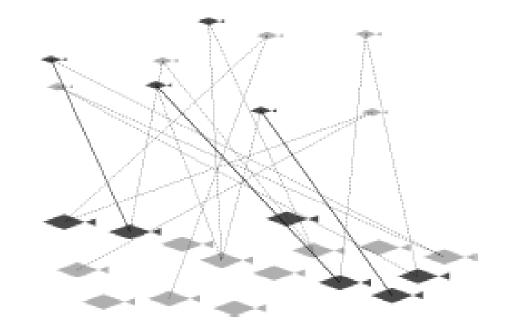


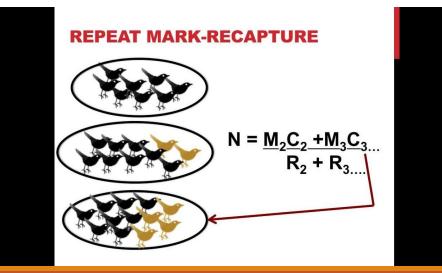
FIG. 1. The simplest form of CKMR. Juveniles are small, adults are big; parents and offspring are linked by lines; dark means sampled, light means unsampled.

Close-kin mark-recapture

- In principle close-kin genetic methods could provide estimates of time-series of (absolute) abundance but there is so much more:
 - estimates of natural mortality; and
 - estimates of fecundity.
- But:
 - The estimates of abundance are not independent (as is always the case for tagging data)
 - Ideally the data themselves should be included in the likelihood function (and not summary statistics) [Mark Bravington says "this does involve much more than some algebra and a few loops"***]

Mark-recapture generally

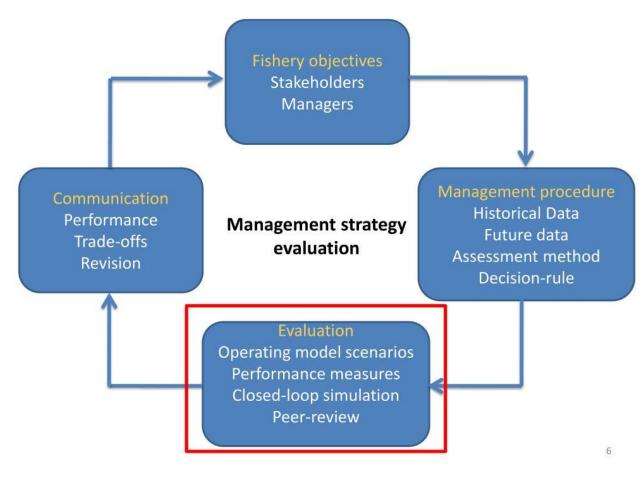
- Mark (or tag) recapture data provide information on:
 - growth rates;
 - Fecundity-at-age;
 - fishing (and natural) mortality; and
 - movement rates.
- There is a need to unify the various approaches taken to include tagging (MULTIFAN seems the most general at present).





Simulation testing and MSE

- Simulation testing is essential to any new package and has led to an increased understanding of the performance of assessment methods.
- MSE is considered state-of the art for evaluating management strategies.
- But how much should a new package include generation of pseudo data sets and closed loop simulations.



New packages **should** include generation of data sets and the capability for closed loop simulations but these should be insufficient for a full evaluation.

Simulation testing and MSE

- Simulation testing should based on data sets generated using the assessment package ("bootstrap feature") which makes testing "easy" but can discourage evaluation of model misspecification.
- Ideally, the ability call "child processes" from within the package will allow analysts to based their simulation analyses / MSEs on approaches (estimation methods and management strategies) coded in languages other than that on which the package is based. Doug Butterworth and his group have done something like this (and it is a core part of Atlantis).
- Development of a common format for simulated data (among packages) sets should facilitate testing of approaches when the operating models are based on different packages.



Management strategy evaluation: best practices

André E Punt^{1,2}, Doug S Butterworth³, Carryn L de Moor³, José A A De Oliveira⁴ & Malcolm Haddon²

¹School of Aquatic and Hshery Sciences, University of Washington, Seattle, WA, 98195, USA: ²CSIRO Oceans and Atmosphere, GPO Box 1538, Hobart, TAS, 7001, Australia: ³Marine Resource Assessment and Management Group (MARAM), Department of Mathematics and Applied Mathematics, University of Cape Town, Rondebosch, 7701, South Africa: ⁴CIFAS Lowestofi Laboratory, Pakefield Road, Lowestofi, Suffolk, NR33 OHT, UK

Abstract

Management strategy evaluation (MSE) involves using simulation to compare the relative effectiveness for achieving management objectives of different combinations of data collection schemes, methods of analysis and subsequent processes leading to management actions. MSE can be used to identify a 'best' management strategy among a set of candidate strategies, or to determine how well an existing strategy performs. The ability of MSE to facilitate fisheries management achieving its aims depends on how well uncertainty is represented, and how effectively the results of simulations are summarized and presented to the decision-makers. Key challenges for effective use of MSE therefore include characterizing objectives and uncertainty, assigning plausibility ranks to the trials considered, and working with decisionmakers to interpret and implement the results of the MSE. This paper explores how MSEs are conducted and characterizes current 'best practice' guidelines, while also indicating whether and how these best practices were applied to two casestudies: the Bering-Chukchi-Beaufort Seas bowhead whales (Balaena mysticetus; Balaenidae) and the northern subpopulation of Pacific sardine (Sardinops sagax caerulea; Clupeidae).

Correspondence: André E Punt, School

of Aquate and Fishery Sdences, University of Washington, Beastle, WA 98195, USA Td.:+1 (206) 221-6319 Pax:+1 (206) 685-7471 B-mail: aquat@uw. edu

Received 26 Jun 2014 Accepted 22 Oct 2014

Keywords Fisheries management, management procedure, management strategy evaluation, simulation, stakeholders, uncertainty

Introduction	304
MSE - the basics	306
Overview of the case-studies	
Bering-Chukchi-Beaufort Seas bowhead whales	
Northern subpopulation of Pacific sardine	
Best practices for MSE	
Establishing objectives and performance statistics	
Selection of uncertainties to consider and selection of operating model parameter values	
Identification of candidate management strategies which could realistically be considered for implementation	
Simulation of application of each management strategy for each operating model	
Presentation of results and selection of a management strategy	
Did the case-studies follow 'best practice'?	
Final comments	

© 2014 John Wiley & Sons Ltd

23

Minimum standards for MSE

- The uncertainties typically included in an MSE:
 - Process error
 - Observation error
 - Model error
 - Errors when conducting assessments
 - Implementation error
- Avoid
 - MSE-lite (adding error to the true operating quantities, even is correlated)

Table 3 List of factors, whose uncertainty commonly has a large impact on management strategy performance, which should be considered for inclusion in any management strategy evaluation.

Productivity

- Form and parameters of the stock-recruitment relationship.
- Presence of depensation.
- Extent of variation and correlations in recruitment about the stock-recruitment relationship.
- Occasional catastrophic montality or recruitment events.

Data-related issues

- CVs and effective samples sizes of data.
- Changes in the relationship between catchability and abundance.
- Changes in survey bias (fishery-independent data).
- Survey and sampling frequency.
- Ageing error.
- Historical catch inaccuracy (bias).
- Outcome (Implementation) uncertainty
- Decision-makers adjust or ignore management advice.
- Realized catches differ from total allowable catches due to misreporting, black market catches, discards, etc.

Non-stationarity

- Changes in the stock-recruitment relationship.
- Time-varying natural mortality (potentially a multispecies operating model).
- Time-varying carrying capacity (regime shift; linked to environmental variables or multispecies effects).
- Time-varying growth and selectivity.
 Other factors
- Spatial and stock structure.
- Technical interactions.
- Time-varying selectivity, movement and growth.
- Initial stock size (unless it is estimated reliably when conditioning the operating model).

Punt et al. (2016)



The next-gen model should be coded in a language such as TMB and be fully open source.

Stock Synthesis Version 1.0

Lets get technical?

Most assessment models are written in ADMB. This is welldeveloped and understood technology. However, there are some major limitations:

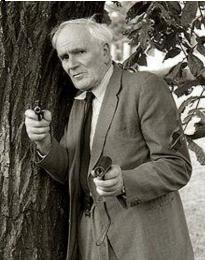
- Dealing with random effects is still a "research topic"
- Effective sampling from a Bayesian posterior is impossible for many models.

The core question for the Next-Gen model: dealing with N

- We want multiple dimensions (partitions) but only the ones we will use:
 - North Pacific crab where females molt annually and males less frequently and we model old and new shell crab leading to an empty partition (old shell females).
- Extending a model to include a new type of partition (i.e. N[y,s,a] >- N[y,s,a,b,q]) is the most painful change to any package because N "links" all model components.



Another key decision: should these chaps change with time (hint: they did; who are they now).



Even more issues

- Harvest control rules that can be specified generically and are fully integrated into the assessment package. The reference point / forecast components of most assessment packages are very case- (location) specific. This has led to (for example) to the need to develop standalone projection components (such as the "West coast rebuilder" for SS – but that has not kept up with SS development or has it?). Ultimately, multi-species, multi-stock models can allow for control rules to evaluate multispecies MSY (or pretty good yield).
- More general (e.g. temporally-correlated) likelihood functions such as a multivariate normal for compositional data and likelihood for correlated indices [e.g. from close-kin MR] (also an easy way to add more likelihood functions, cf. CASAL/CASAL2)
- Ability to parameterize the model using leading parameters (MSY, B_{MSY}/K, etc.)

Even more issues (Continued)

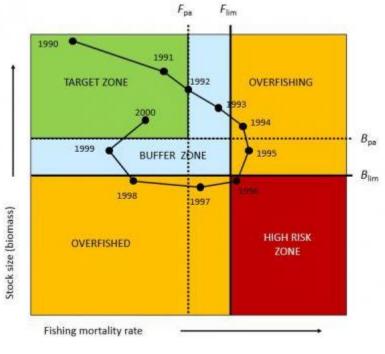
- Making variance estimation using MCMC (much) faster (ADMB vs CASAL?) Right now it is almost impossible to develop a Bayesian posterior within an assessment review meeting.
- When to dispense with areas-as-fleets as the recommended default? We know that areas-as-fleets leads to biased results but the alternative is spatial structured models and they require a lot of data (and are complicated)
- Making use of random effects on selectivity easier to use. At present the "VPA approach" to selectivity is much less used than is probably appropriate (but what to do in data poor cases).



Slideshare.net

Even more issues (Continued)

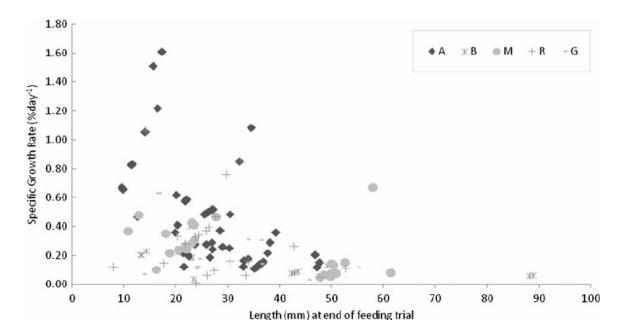
- Dynamic reference points Most of packages do not model time-varying parameters, particularly in the projection phase. However, the need for "current" reference points continues to increase.
- Research on parameterizing stock assessment models (e.g. facilitate model convergence in maximum likelihood and Bayesian contexts).
- **Continued "co-evolution" on diagnostic statistics** and development of "best practices" (or at least "common things to think about").
- **Stock-recruitment relationships.** Where are we with non-parametric approaches in integrated assessments.
- **Partitions**. How many partitions and can the number of partitions be changed without recoding?



MRAG

Yet more issues (Continued)

- What about size-structured models-I: Most "packages" for size-structured models are for "data poor" cases (e.g. LIME), with "integrated" size-structured models mainly bespoke models (with the exception of CASAL/CASAL2). Is this because size-structured problems are unique (or just the assessment scientists concerned eclectic)?
- What about size-structured models-II: Size-structured models are applied to "hard-to-age" animals such as prawns, and rock lobsters. Some of these species (e.g. abalone) are characterized by extreme variation in biological parameters spatially. Do we know what to do in these cases?



Is this the model you want?



Because you may end up with this



Which packages make the final and who are the winners?

- Multiple "stocks" and species in one assessment.
- Spatial structure
- Age-length models
- Tagging data

Stock structure: GADGET, Poseidon, CASAL

Spatial structure: GADGET, Poseidon, MULTIFAN, CASAL

True length dynamics: SS (Sorta), GADGET, GMACS, CASAL

Tagging data: SS (sorta), GADGET, MULTIFAN, GMACS, CASAL



Which packages make the final and who are the winners?

- Dealing with random effects and data weighting "correctly"
- Biological interactions
- Simulation and Management Strategy Evaluation

Random effects (or Bayesian): SS, ASAP, WHAM (RE), Coleriane, GMACS, SAIGE, JABBA, a4a, CASAL, SAM (RE)

Biological interactions: GADGET, CEATTLE

Close kin general: TBA

Simulation: SS, ASAP, BAM, GADGET, Poseidon, WHAM, MULTIFAN, a4a, CASAL

MSE: SS (really?), BAM, GADGET, MULTIFAN, a4a



• Stock Synthesis

 Many options but access to code is restricted; random effects only as Bayesian; source code..

GADGET

 Many options; limited user data (at present); slooow

• CASAL

 Many options (incl. stock structure); efficient code structure ; use of MDL; random effects only as Bayesian; is it developed?

MULTIFAN

 Strong spatial structure options; quite slow; source code?



Which model is which animals

Final thoughts-I

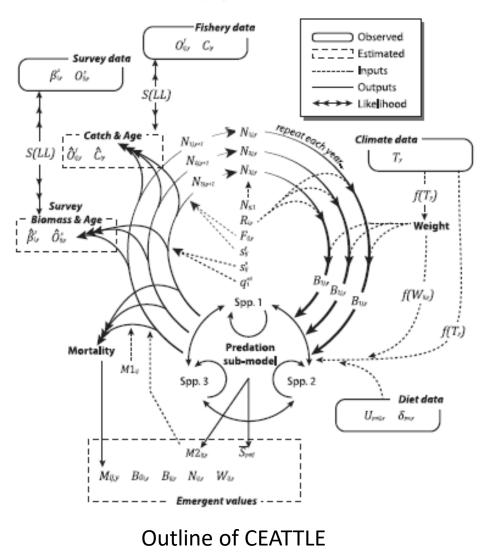
- How to ensure any packages are:
 - well maintained and documented (with test data sets);
 - efficient (in the sense of not being unduly complicated by having to provide specifications for features that don't matter or no-one uses);
 - who controls the features (and are the gatekeepers of development); and
 - what about automatic template for "standard assessments".



Clipartpanda.com

Final thoughts-II

- What about multi-species considerations
 - GADGET allows for both technical and biological interactions (but does not fit to diet data); what about methods such as CEATTLE?
 - Several MICE models have been developed but they are case-specific by design. Perhaps this is an insolvable problem?
 - How many stocks do we have sufficient data to apply models with dynamic predation and competition?





Often it is not the lack of features that is the problem but rather a lack of training. Enter CAPAM





Identify a giant and what they are known for (not Rick, Ray, Mark or Kelli et al) and you can ask a question!

