https://doi.org/10.17895/ices.pub.8009

Strategic Initiative on Stock Assessment Methods (SISAM)

BOSTON

WCSAM
2013

WORLD CONFERENCE ON
STOCK ASSESSMENT METHODS

WCSAM Workshop 15-16 July 2013, Boston USA

SISAM Leadership

- SISAM Co-Chairs:
  - Mark Dickey-Collas
  - Steve Cadrin
- SISAM Steering Committee:
  - Doug Butterworth
  - Richard Methot
  - Carmen Fernandez
  - José De Oliveira
- WCSAM Steering Committee:
  - Jean-Jacques Maguire
  - Ernesto Jardim
  - Laurie Kell,
  - Cathy Dichmont
  - Ana Parma
  - Victor Restrepo
  - Yimin Ye

SISAM Objectives

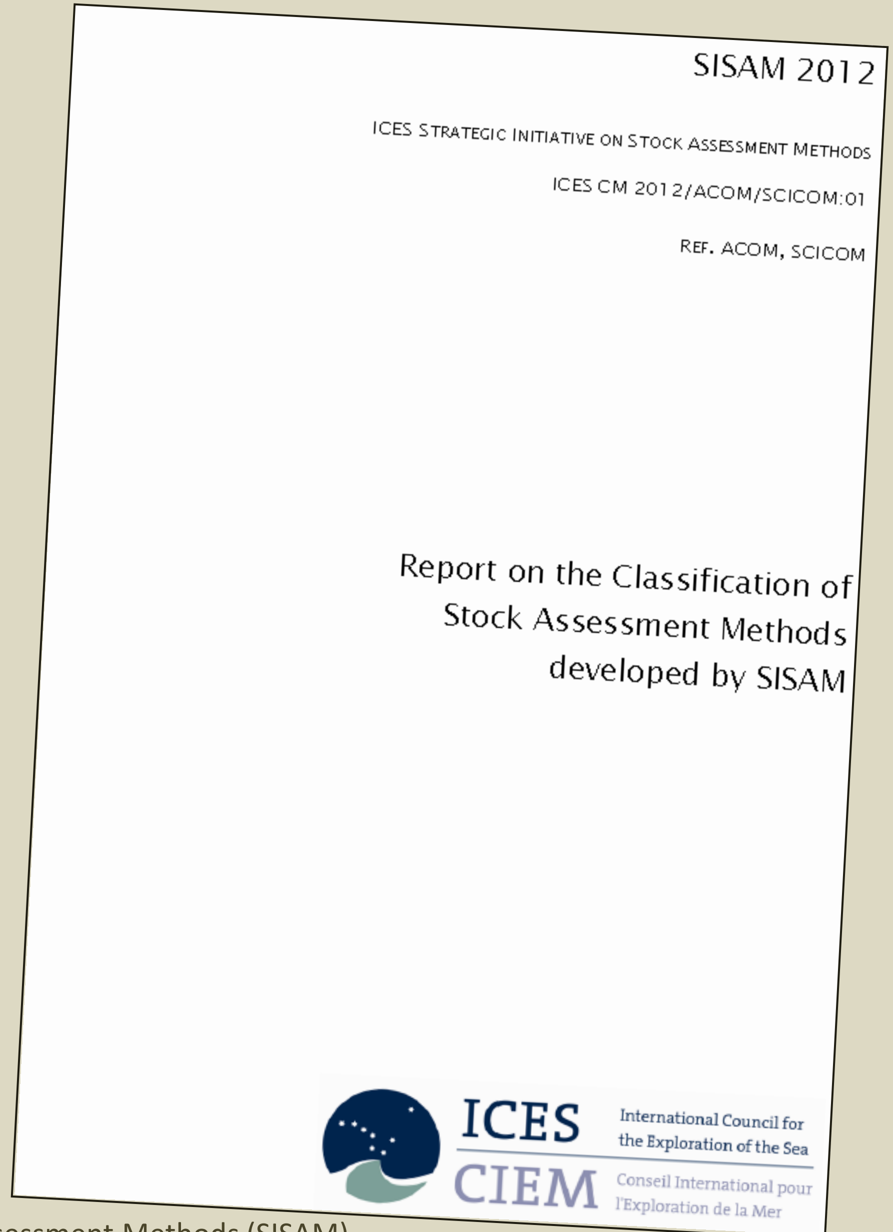
- The **Strategic Initiative for Stock Assessment Methods** (SISAM) is designed to assure that ICES scientists can apply the best methods when developing management advice.
- Other Regional Fishery Management Organizations and national fishery organizations have a similar goal, so success of SISAM will have benefits for the entire international fishery science community.

SISAM Process

1. identification of the current set of available methods;
2. guidance in the selection of the most appropriate methods for a particular application;
3. education and access to expert information regarding method usage;
4. encouragement for further testing and development of methods to more closely align with particular management needs and to take advantage of advances in statistical theory, computing power, and new knowledge.

1. Identification of available methods

- Classification scheme:
  - Catch only
  - Time series models
  - Biomass dynamics models
  - Delay-difference models
  - Age-structured production models
  - Virtual Population Analysis
  - Statistical catch-at-age
  - Integrated Analysis models
- Questionnaire to RFMOs


SISAM 2012

ICES Strategic Initiative on Stock Assessment Methods

ICES CM 2012/ACOM/SCICOM:01

Ref. ACOM, SCICOM

Report on the Classification of Stock Assessment Methods developed by SISAM

ICES International Council for the Exploration of the Sea

CIEM Conseil International pour l'Exploration de la Mer

2. Guidance on Most Appropriate Methods

- Simulation-based process to evaluate performance of alternative methods for meeting the needs of fishery management.
  1. ICES Methods Working Group (José De Oliveira, Anders Nielsen, Rebecca Rademeyer and Doug Butterworth, …)
  2. NOAA Contract to use POPSIM (Jon Deroba, Jessica Blaylock, Murali Mood, Chris Legault, Pal Rago, …)
  3. CEFAS Contract with alternative treatment of measurement and process errors (José De Oliveira and Tim Earl)
  4. WCSAM workshop to present and consider results for general conclusions and recommendations

Contributing ‘Model Experts’

- Leire Ibaibarriaga, Erik Williams , Grant Thompson , Jon Brodziak, Doug Swain, Chris Legault, Jim Ianelli, Paul Spencer, Canales, Alexadre Silva, Bertignac, Erik Williams , Anders Nielsen, Nils Hintzen, Valero, Brites Azevedo, Alberto Murta, Jon Deroba, De Moor, Jose De Oliveira, Pete Hulson, Dana Hanselman, Rebecca Rademeyer and Doug Butterworth
- Workshop helpers: Owen Nichols, Daniel Goethel, Ben Galuardi, Jannica Haldin

3. Education on Method Usage

- ICES Symposium: **The World Conference on Stock Assessment Methods for Sustainable Fisheries** (Boston, USA, 17-19 July 2013).

BOSTON
WCSAM
2013
WORLD CONFERENCE ON
STOCK ASSESSMENT METHODS

4. Further Model Development

- SISAM is the first step in a long-term strategic initiative.
- Results and publications from the workshop and conference will encourage further testing and development of methods to more closely align with particular management needs and to take advantage of advances in statistical theory, computing power, and new knowledge.

METHODS TESTING:

THE DESIGN OF SIMULATION EXERCISES

Doug Butterworth

MARAM (Marine Resource Assessment and Management Group)
Department of Mathematics and Applied Mathematics
University of Cape Town, Rondebosch 7701, South Africa

ICES WORKING GROUP ON METHODS OF FISH STOCK ASSESSMENTS

LISBON OCTOBER 2012

TOR

- a) Assemble 10–12 datasets from ICES that characterize the breadth of life-history strategy, data quality, population dynamics, and assessment problems.
- b) Prepare a publication (to be presented to the SISAM symposium), using these datasets, that explores providing guidelines on simulation testing of assessment models.

TOR a)

STOCKS SELECTED

North Sea cod
North Sea plaice
North Sea herring
North Sea haddock
Northern hake
Spurdog
Biscay anchovy

Iberian sardine
Southern horse mackerel
N Atlantic albacore tuna
US W coast canary rockfish
G Bank yellowtail flounder
South African anchovy

TOR b)

SIMULATION

Discussion centred on the development of an assessment comparison and simulation testing framework

PROPOSED SISAM WORKSHOP SCHEME FOR CHOSEN DATA SETS

- I. Different models, fixed settings
- II. Diagnostics and optimised settings
- III. Simulations: observation error only

  (a) self test (b) cross test
- IV. Simulations: observation + process error
- V. Simulations: Grand questions

  May need to force more contrast in data

MODEL FITS TO REAL DATA SETS

For key assessment outputs – how dependent on method (model) chosen?

Try many models

Simple to complex continuum

x

- I. Different models, fixed settings
- II. Diagnostics and optimised settings

  AIC, cross-validation, etc.

EXTENSION TO SIMULATION

Difficulty with approaches used previously

Generic – so does result apply to MY stock?

Thus investigate for actual stocks

Base on Management Procedure (MSE) testing protocol developed in IWC

Key consideration – robustness to uncertainty

Consider alternative plausible scenarios (assessments) which MUST be consistent with available data

Apply the "CONDITIONING" concept

CONDITIONING SIMULATIONS

Each pseudo dataset is generated from what could be the real underlying dynamics for the stock concerned (as provided by a plausible assessment model), with errors added consistent with the error distributions as estimated in that assessment

TWO TEST TYPES: SELF/CROSS

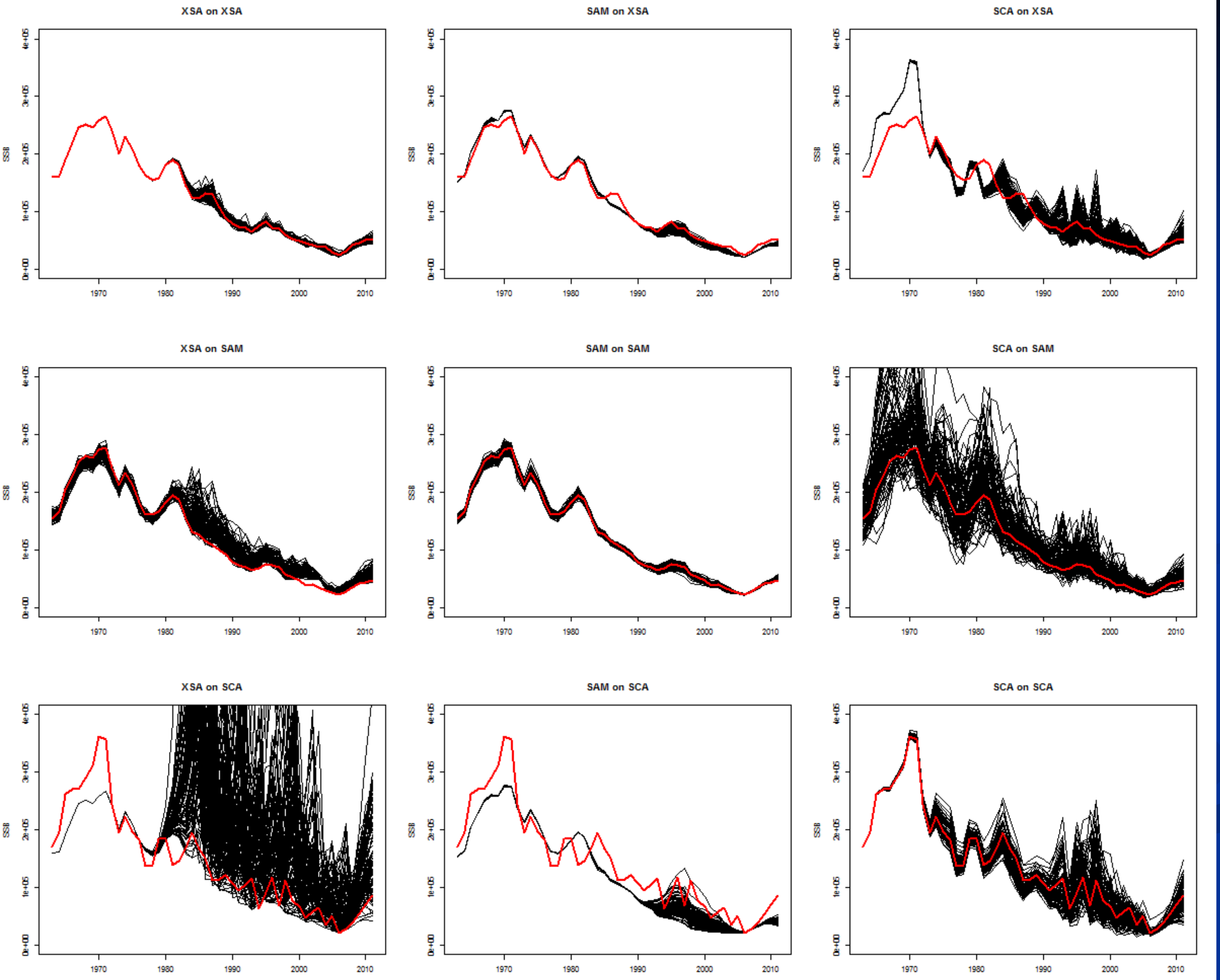
PERFORMANCE COMPARISON PLOT

Rows : "Truth" as provided by a model

Columns: Estimates from the model applied to pseudo-data

Cell contents: Performance statistic, here SSB

[Most pertinent would be the catch under the intended harvest strategy]



XSA on XSA
SAM on XSA
SCA on XSA
XSA on SAM
SAM on SAM
SCA on SAM
XSA on SCA
SAM on SCA
SCA on SCA
SSB
0e+00
1e+05
2e+05
3e+05
4e+05
1970
1980
1990
2000
2010

TWO TEST TYPES: SELF/CROSS

PERFORMANCE COMPARISON PLOT

Rows : "Truth" as provided by a model

Columns: Estimates from the model applied to pseudo-data

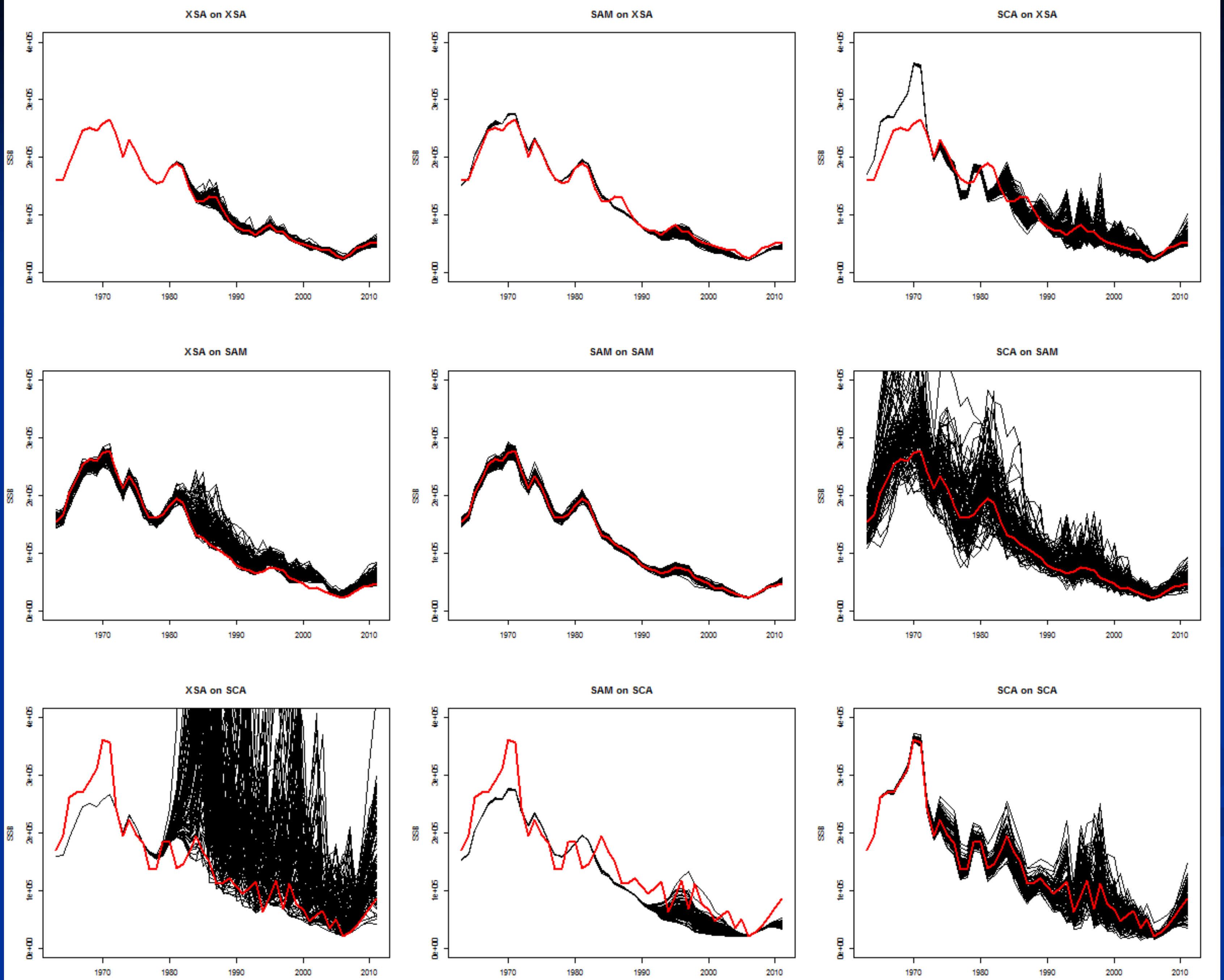
Cell contents: Performance statistic, here SSB

SELF TEST: Diagonals

How well does the model estimate itself

CROSS TEST: Off-diagonals

How well does it estimate other models


XSA on XSA
SAM on XSA
SCA on XSA
XSA on SAM
SAM on SAM
SCA on SAM
XSA on SCA
SAM on SCA
SCA on SCA
SSB
1970
1980
1990
2000
2010

PROPOSED SISAM WORKSHOP SCHEME FOR CHOSEN DATA SETS

III. Simulations: Observation Error only

Simulated randomness only in data generated

Underlying dynamics unchanged over simulations

"EASY" to implement

BUT Catch … - observation or process error?

PROPOSED SISAM WORKSHOP SCHEME FOR CHOSEN DATA SETS

IV. Simulations: Observation + Process Error

- Simulated randomness now also in processes such as recruitment
- Underlying dynamics changes over simulations

"DIFFICULT" to implement

- Can't simply generate alternative recruitment residuals, as actual catches couldn't be taken in some cases
- Generate residuals from parameter variance-covariance matrix to accommodate correlations implied

WHICH WAY TO SIMULATE?

Difficulty with approaches used previously

Generic – so does result apply to MY stock?

Case-specific conditioning – results apply to MY stock – but can anything be said about other stocks, or any generic inference drawn?

Approach?

Repeat for many stocks to see whether patterns emerge which might justifiably be considered reliable general inferences

PROPOSED SISAM WORKSHOP SCHEME FOR CHOSEN DATA SETS

- I. Different models, fixed settings
- II. Diagnostics and optimised settings
- III. Simulations: observation error only

  (a) self test (b) cross test
- IV. Simulations: observation + process error
- V. Simulations: Grand questions

  May need to force more contrast in data

GRAND QUESTIONS

Examples:

- How important is it to have good and frequent age data?
- Does VPA's assumption of catch-at-age being exact matter?

What is the best approach to simulation testing to address this?

Is conditioning on real datasets appropriate – more contrast needed for effective discrimination?

Application of POPSIM – Jon Deroba

Thank you for your attention

With acknowledgements to other participants in the ICES Methods Working Group who assisted in developing this framework

Simulation Work

PopSim and Plots

Jonathan J. Deroba

July 15, 2013

WCSAM Workshop, Boston, MA

PopSim

Prepackaged data generation software

Give PopSim:

- biological characteristics (e.g., M, wts, maturity)
- F, selectivity, recruits, initial N, q (i.e., from model)
- CV of noise in catch and survey (based on residuals)
- effective sample size for age comps

Generates survey obs, catch obs, and age comps based on above

Survey and catch lognormal, age comp multinomial

See interface

PopSim

For future consideration

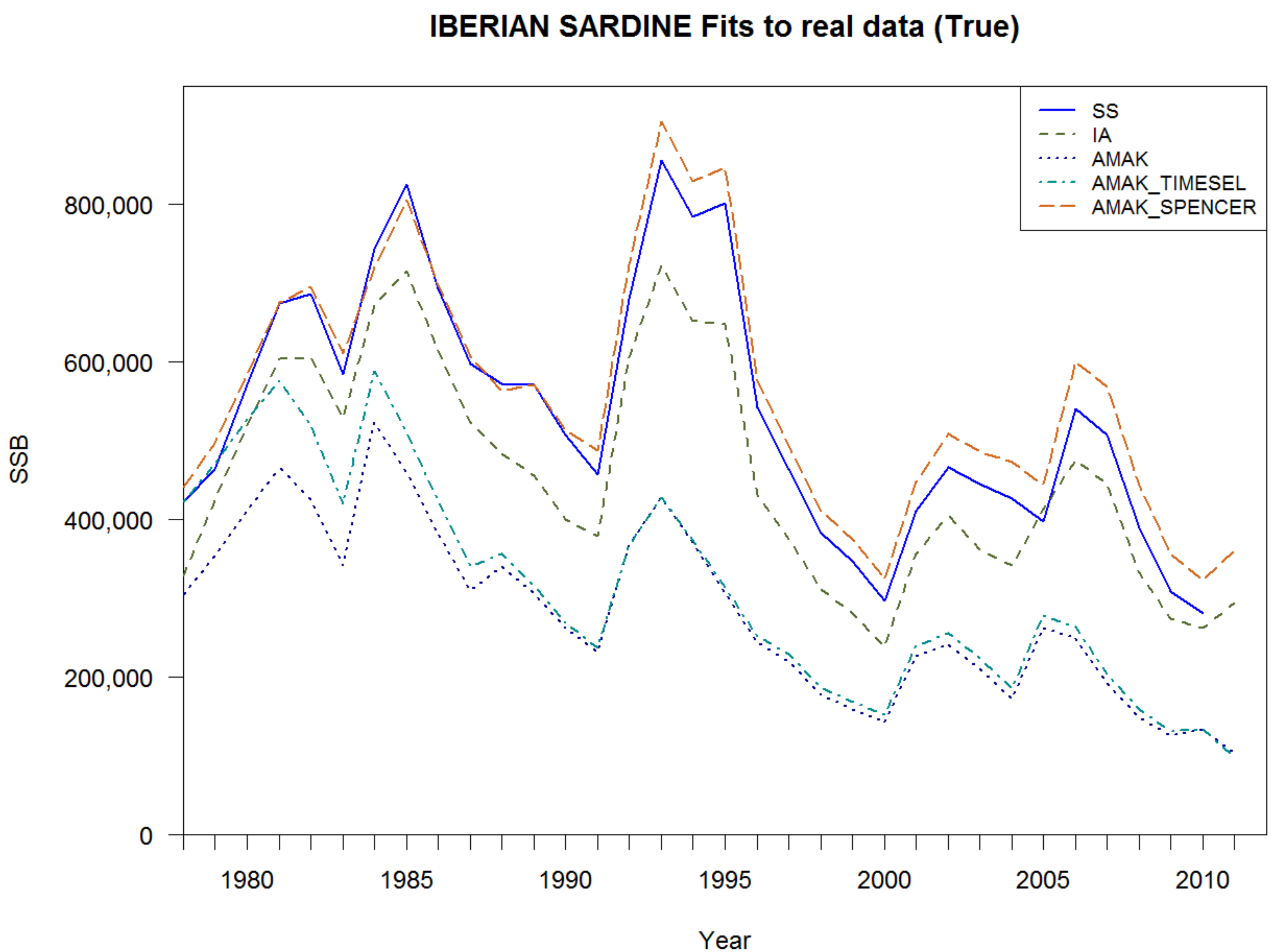
Disadvantages:

- inflexible
- communication issues (e.g., scale, “residual”)

Advantages:

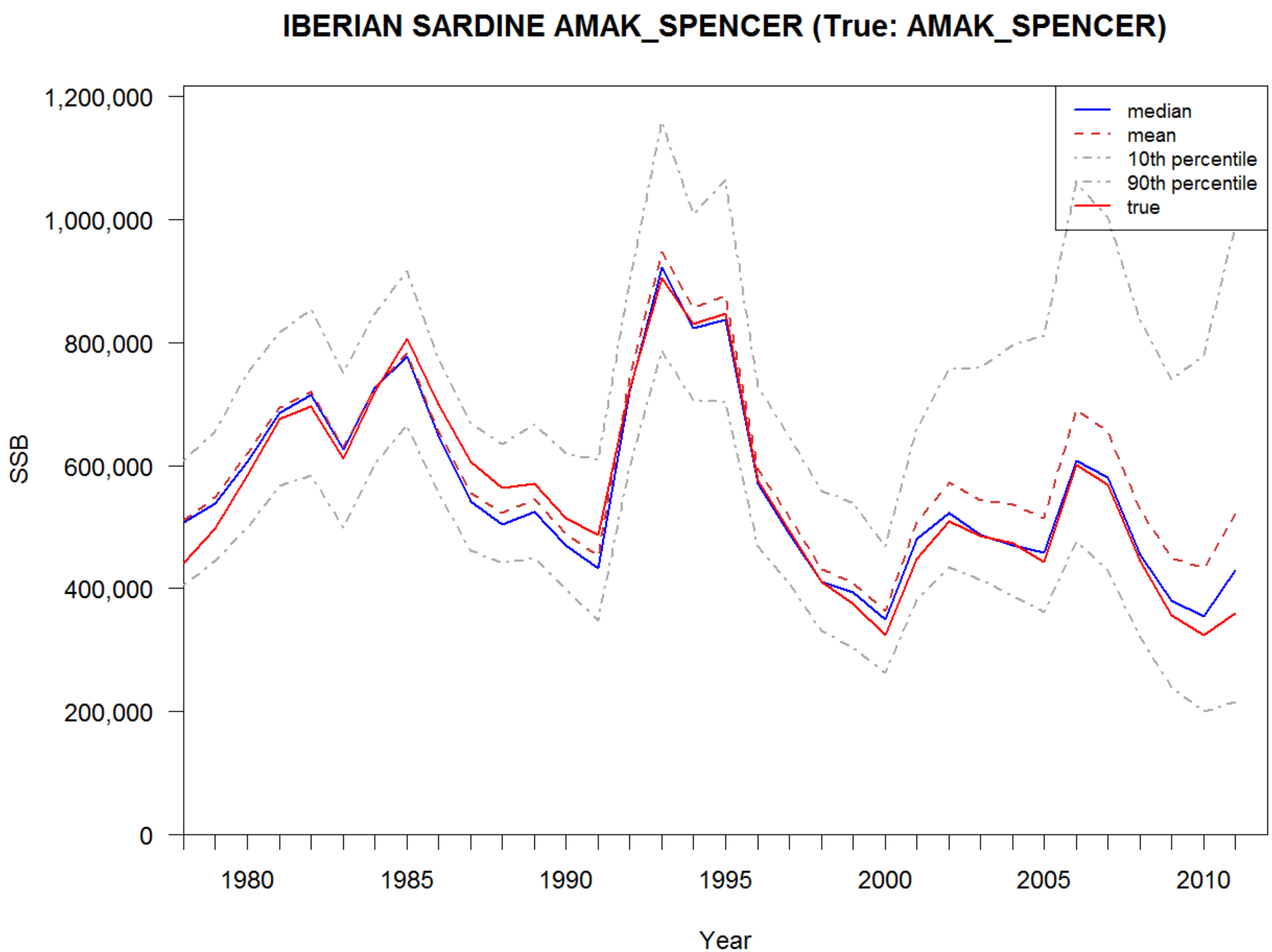
- control and consistency
- no single model advantage

Plots

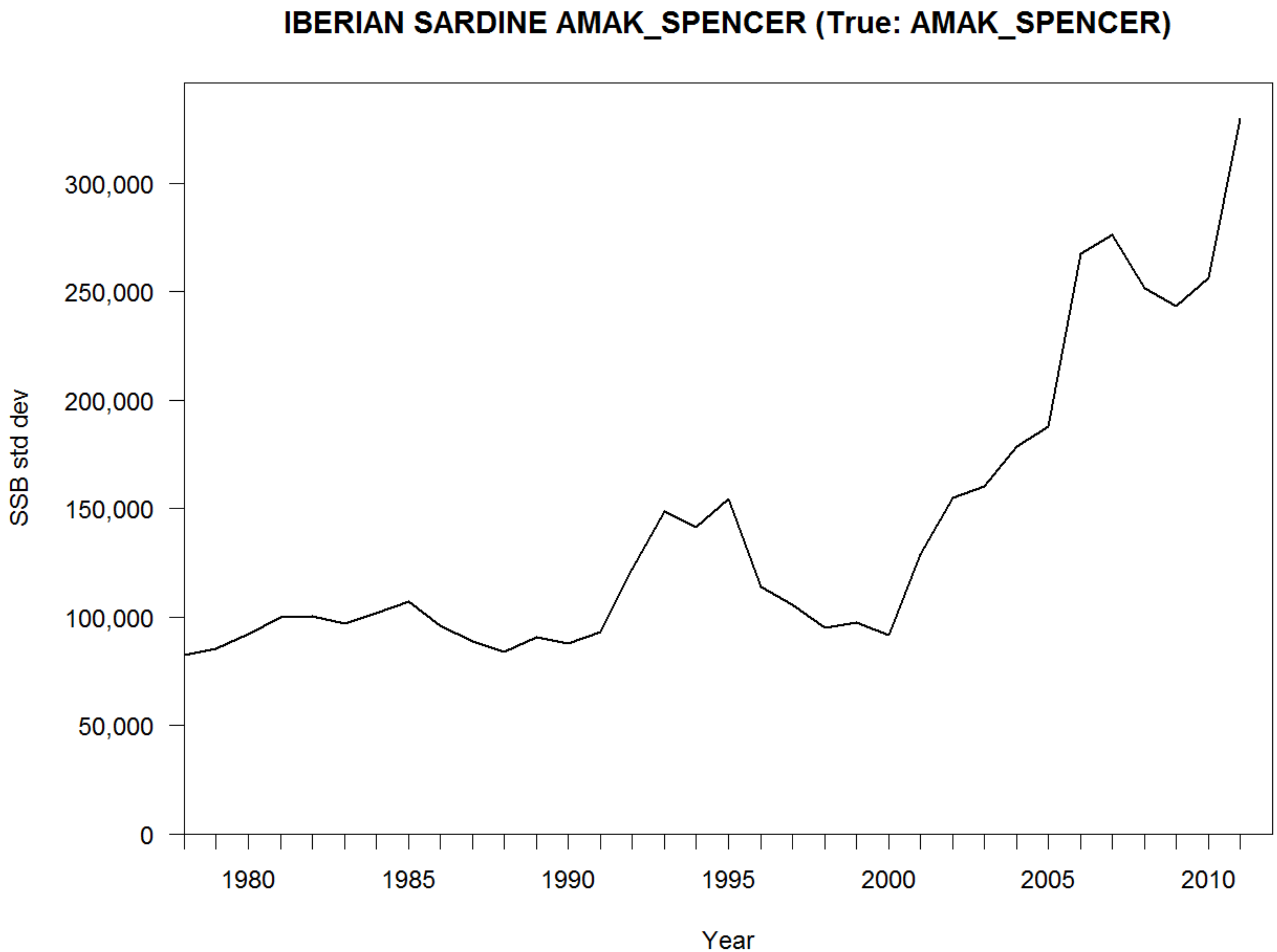


IBERIAN SARDINE Fits to real data (True)
SS
IA
AMAK
AMAK_TIMESEL
AMAK_SPENCER
800,000
600,000
400,000
200,000
0
SSB
1980
1985
1990
1995
2000
2005
2010
Year

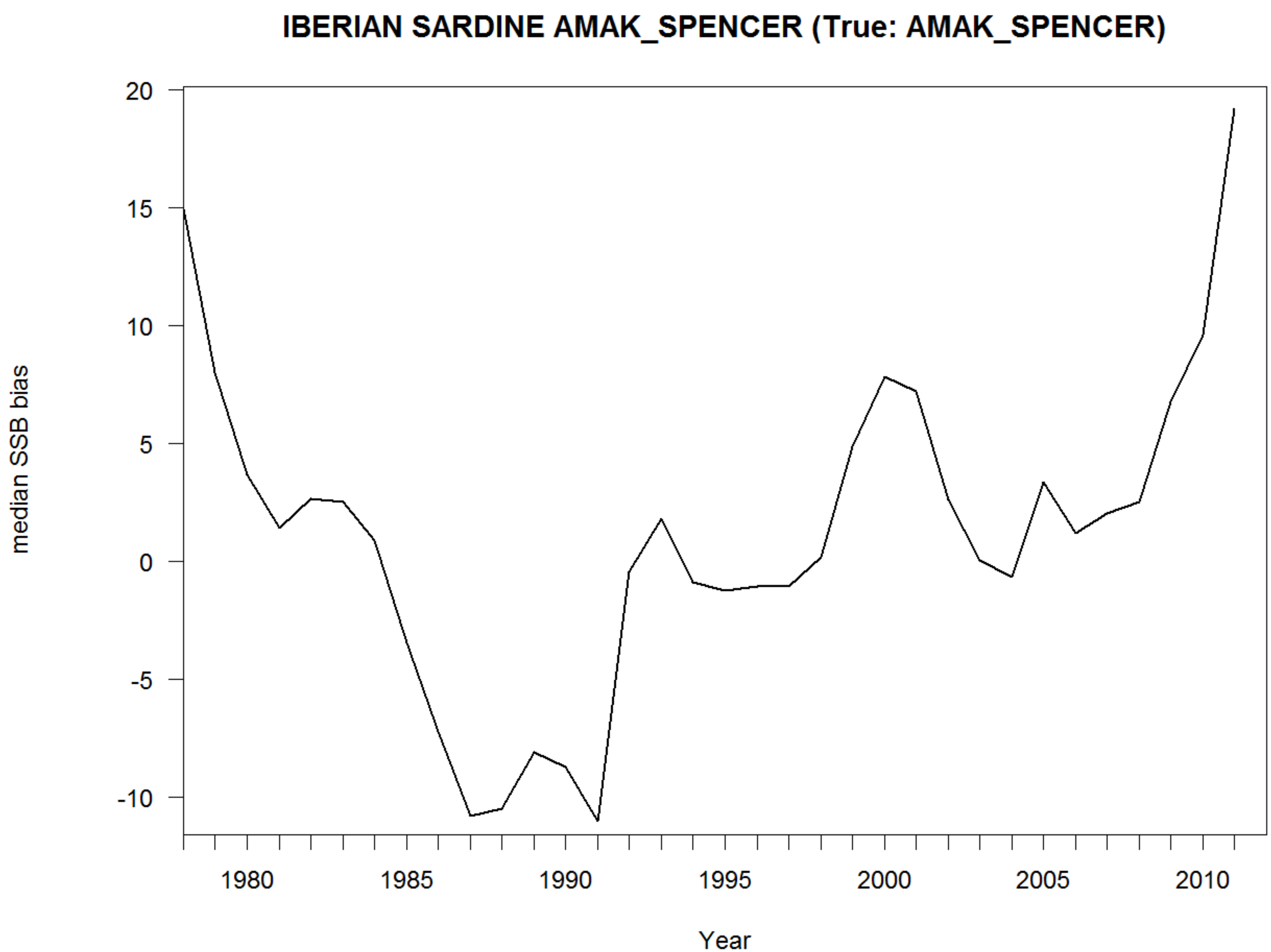
Plots


IBERIAN SARDINE AMAK_SPENCER (True: AMAK_SPENCER)
median
mean
10th percentile
90th percentile
true
SSB
Year
1,200,000
1,000,000
800,000
600,000
400,000
200,000
0
1980
1985
1990
1995
2000
2005
2010

Plots


IBERIAN SARDINE AMAK_SPENCER (True: AMAK_SPENCER)
SSB std dev
300,000
250,000
200,000
150,000
100,000
50,000
0
1980
1985
1990
1995
2000
2005
2010
Year

Plots


IBERIAN SARDINE AMAK_SPENCER (True: AMAK_SPENCER)
median SSB bias
Year
20
15
10
5
0
-5
-10
1980
1985
1990
1995
2000
2005
2010

Plots

Repeat for F

Participation

Species_name	Total
BISCAY ANCHOVY	1
GBYT FLOUNDER	13
IBERIAN SARDINE	5
NORTHERN HAKE	1
NS COD	11
NS HADDOCK	5
NS HERRING	10
NS PLAICE RAISED RECON	1
S H MACKEREL	6
SA ANCHOVY	2
SPURDOG	3
WC CANARY ROCKFISH	2
Grand Total	60

species_name	Total
GBYT FLOUNDER	20
IBERIAN SARDINE	2
NS COD	3
S H MACKEREL	1
SA ANCHOVY	4
Grand Total	30

Georges Bank Yellowtail Flounder WGSAM Summary

Christopher M. Legault
NMFS - NEFSC
WGSAM
Boston July 2013

Thanks to everyone who participated!

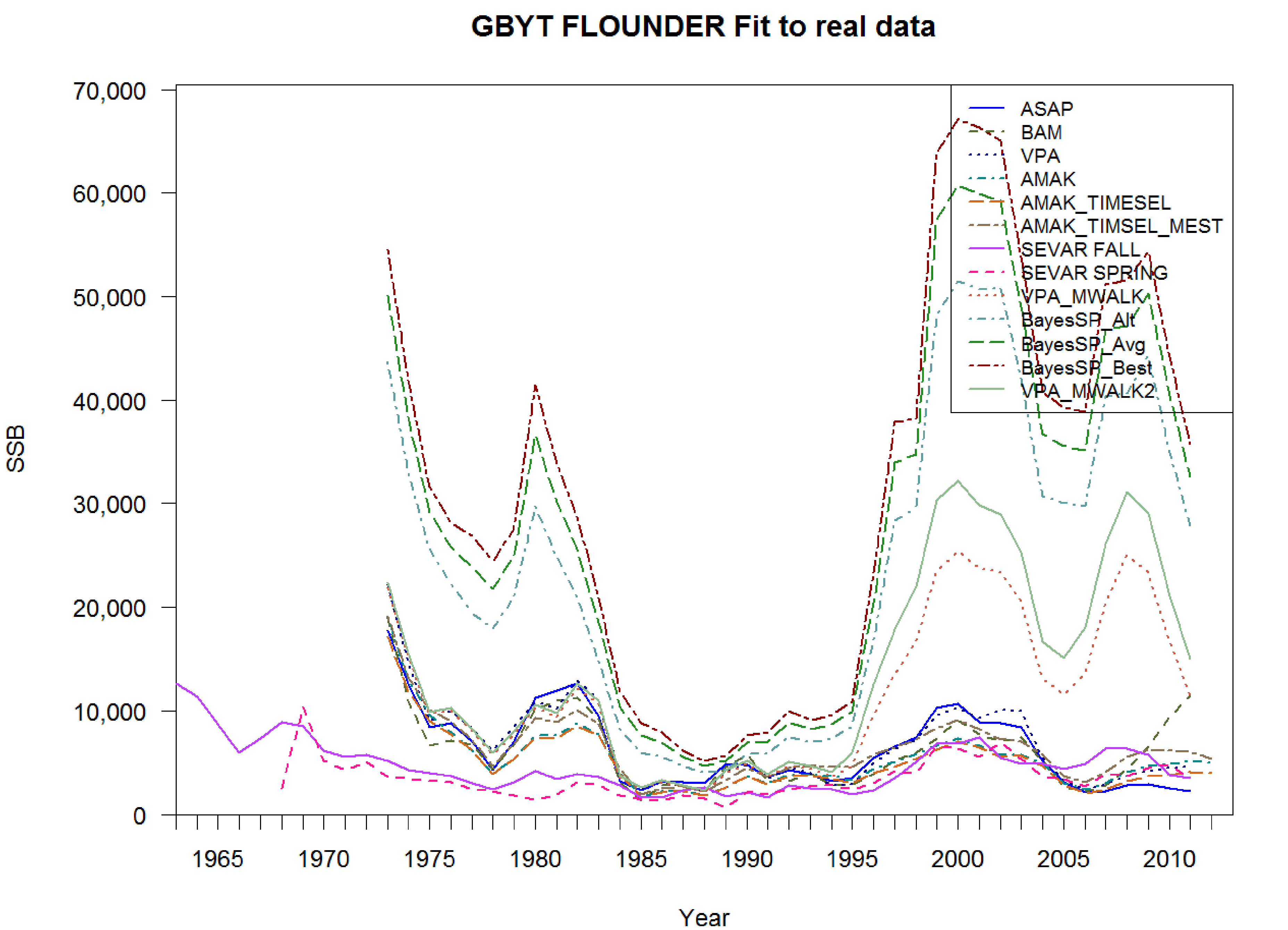
Overview of Stock

- Adapt VPA 1973-2011
  - Ages available 1973 onward
  - Fishery began 1935
  - International fishery and assessment (US and Canada)
  - Surveys began 1963 (Fall), 1968 (Spring), 1987 (DFO), 1982 (Scallop, age-1 only)
- Main feature: Strong retrospective
  - Relative F decreased 1995 but survey Z high throughout
  - Splitting surveys 94/95 worked originally, but not anymore
  - Explored increasing recent catch, increasing recent M, increasing both recent catch and M, as well as retro adjustments
    - All resulted in similar catch advice

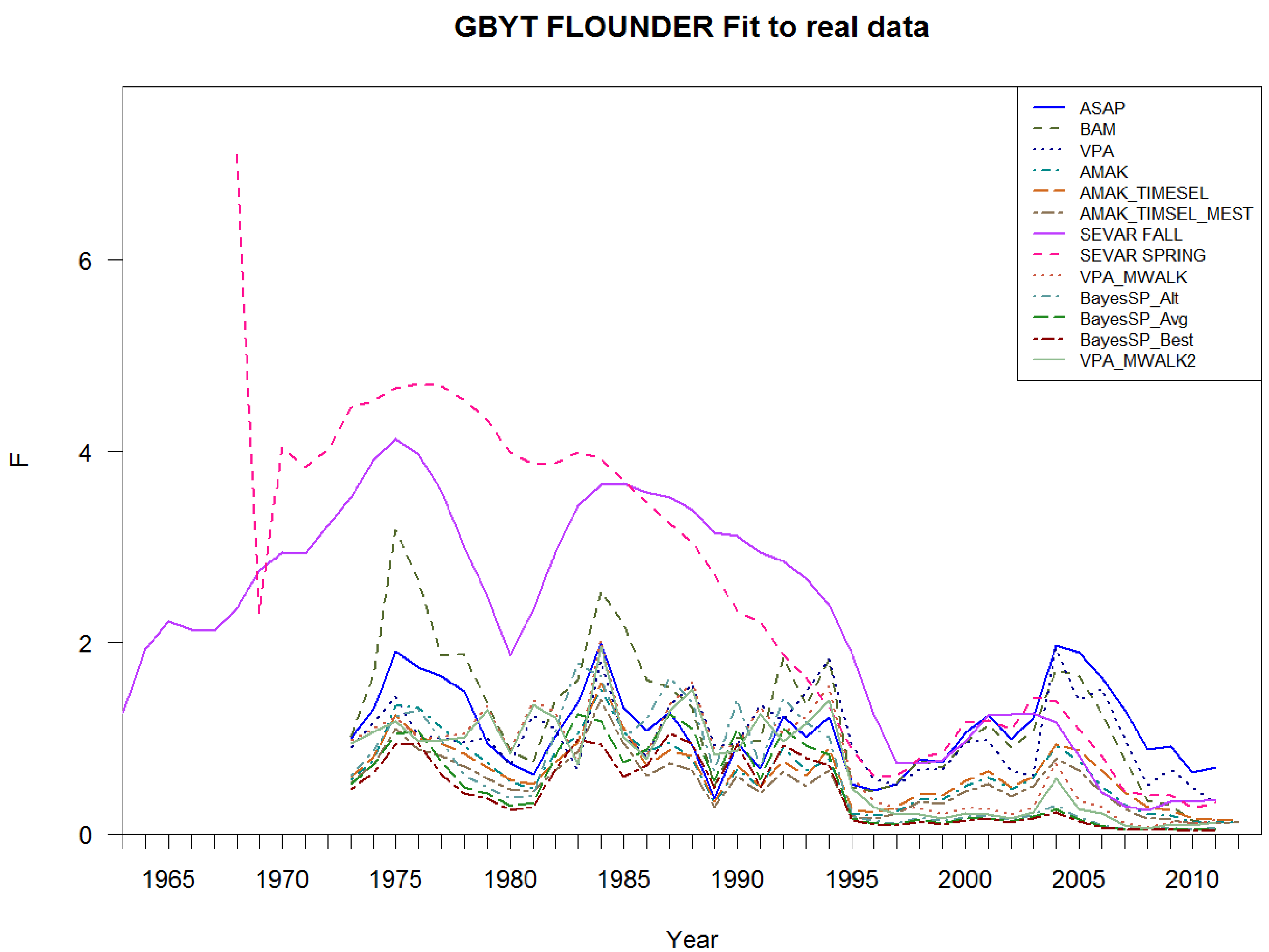
http://www2.mar.dfo-mpo.gc.ca/science/TRAC/rd.html

What was done

- Fits to original data
  - VPA with retrospective adjustment (Legault)
  - ASAP with time varying q (Legault)
  - VPA with time and age varying M 2 runs (Swain)
  - BAM (Williams)
  - AMAK 3 runs including time varying selectivity (Ianelli)
  - SEVAR fall and spring (Thompson)
  - Bayesian Surplus Production Model 3 runs (Brodziak)
  - Time series analysis (Gudmundsson) (not shown)
    - All random elements enormous



GBYT FLOUNDER Fit to real data
SSB
Year
70,000
60,000
50,000
40,000
30,000
20,000
10,000
0
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
ASAP
BAM
VPA
AMAK
AMAK_TIMESEL
AMAK_TIMSEL_MEST
SEVAR FALL
SEVAR SPRING
VPA_MWALK
BayesSP_Alt
BayesSP_Avg
BayesSP_Best
VPA_MWALK2


GBYT FLOUNDER Fit to real data
ASAP
BAM
VPA
AMAK
AMAK_TIMESEL
AMAK_TIMSEL_MEST
SEVAR FALL
SEVAR SPRING
VPA_MWALK
BayesSP_Alt
BayesSP_Avg
BayesSP_Best
VPA_MWALK2
F
0
2
4
6
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
Year

What did we learn?

- Wide range of SSB and F estimates
  - Some real, some artificial
- Trends differ
  - Implications for “story telling” about historical events
- Catch advice huge range -3,000 t to 12,000 t
  - Wide range of control rules
  - Not reported from all models

- External/local information important

Update

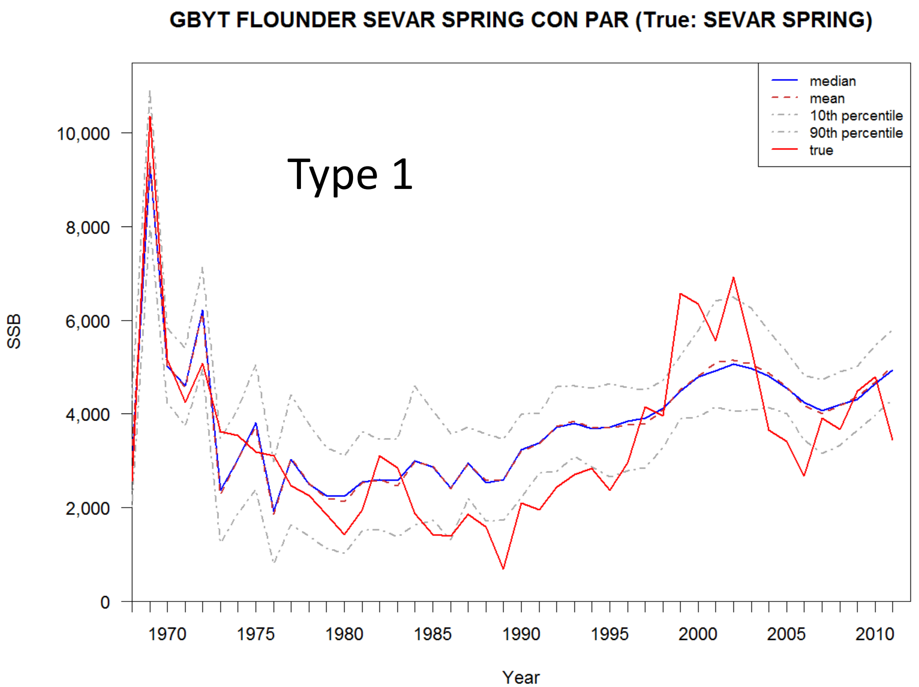
- This assessment updated last month
- Since 2008 both catch and survey have declined
  - Catch in 2012 lowest since 1940 (well below quota)
  - Relative F flat since 2008
  - Survey catch curve Z remains high
  - All surveys indicate poor recent recruitment
- Canadian fishermen could not find yellowtail
- Recommending lowest quota ever

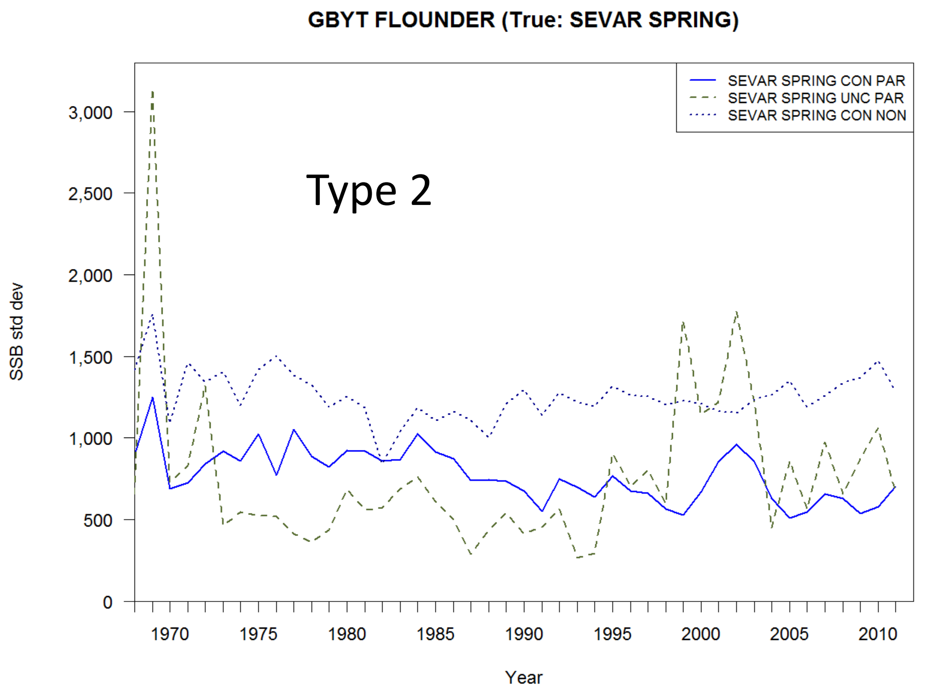
What was done part 2

- Generate 100 data sets and fit models

- Model fit to data (data generator) [20 combos]
  - ASAP with time varying q (ASAP)
  - VPA Mwalk 2 runs (ASAP)
  - VPA with time varying q (ASAP)
  - SEVAR fall and spring (ASAP)
  - VPA Mwalk (VPA Mwalk)
  - VPA Mwalk2 (VPA Mwalk2)
  - SEVAR fall and spring (VPA Mwalk and Mwalk2)
  - SEVAR fall and spring (VPA)
  - SEVAR fall (SEVAR fall 3 types of uncertainty)
  - SEVAR spring (SEVAR spring 3 types of uncertainty)

Not going to show all the figures – see http://www.nefsc.noaa.gov/ExternalDrive/ (username and password are both SISAM) for all the figures that were produced





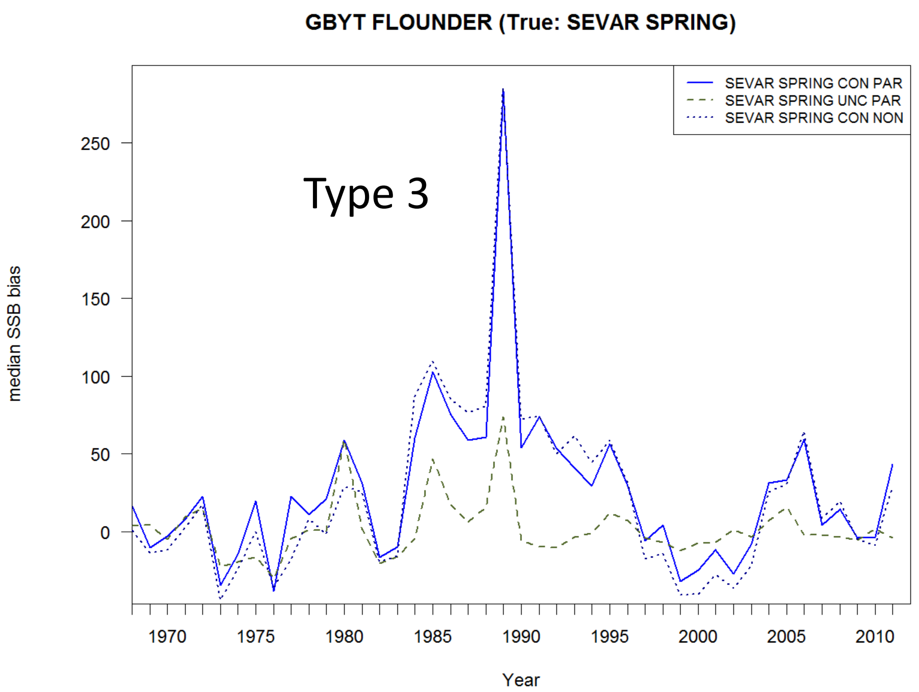
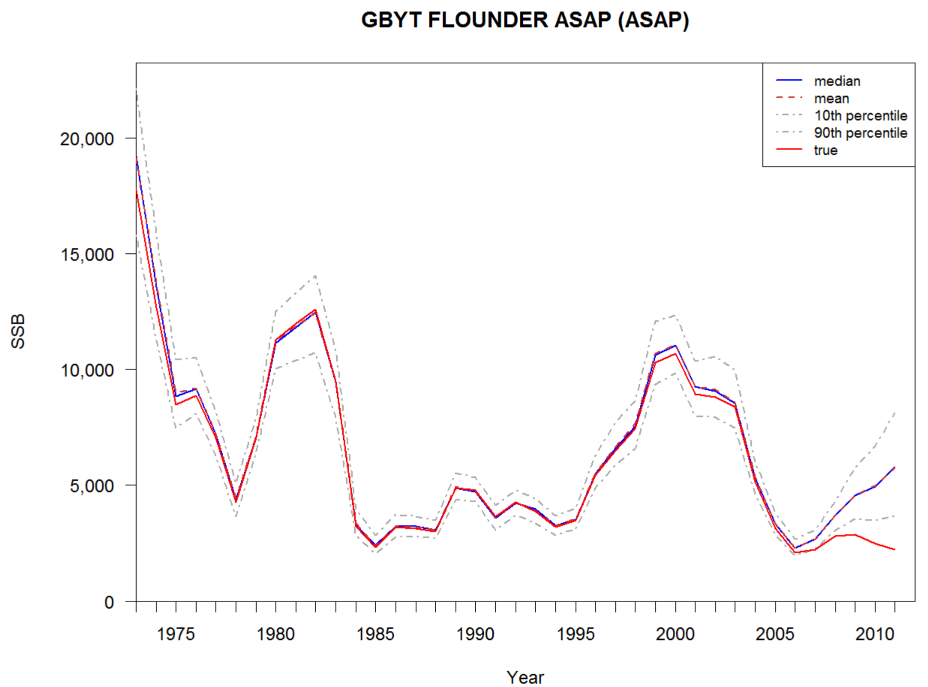


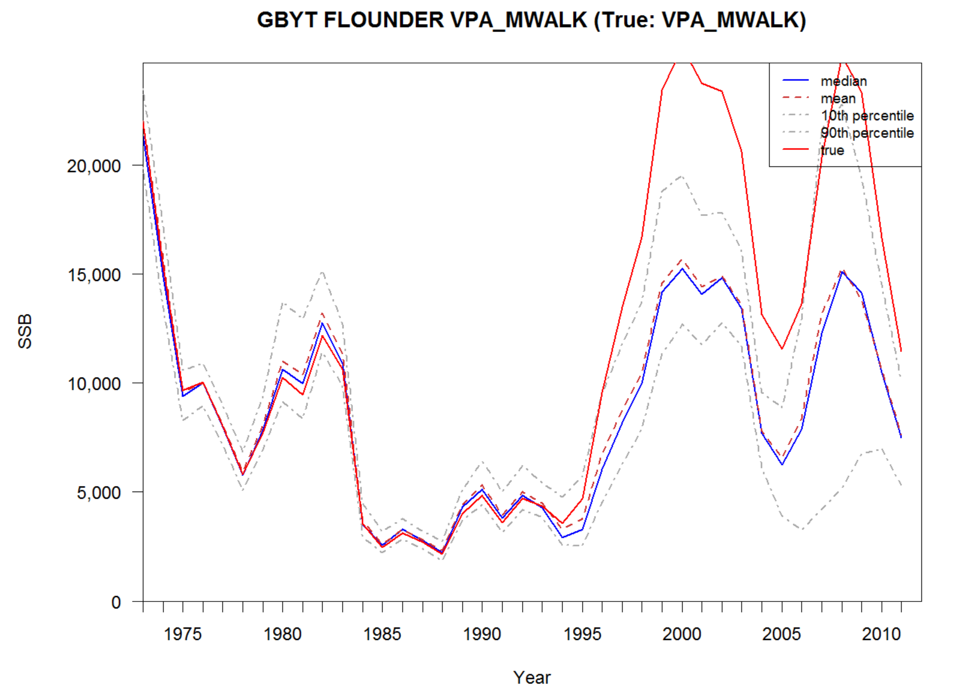
Figure Types Generated

1. True and four summary statistics from 100 replicates (median, mean, 10th and 90th percentiles)
2. Standard deviation from 100 replicates
3. Median bias from 100 replicates

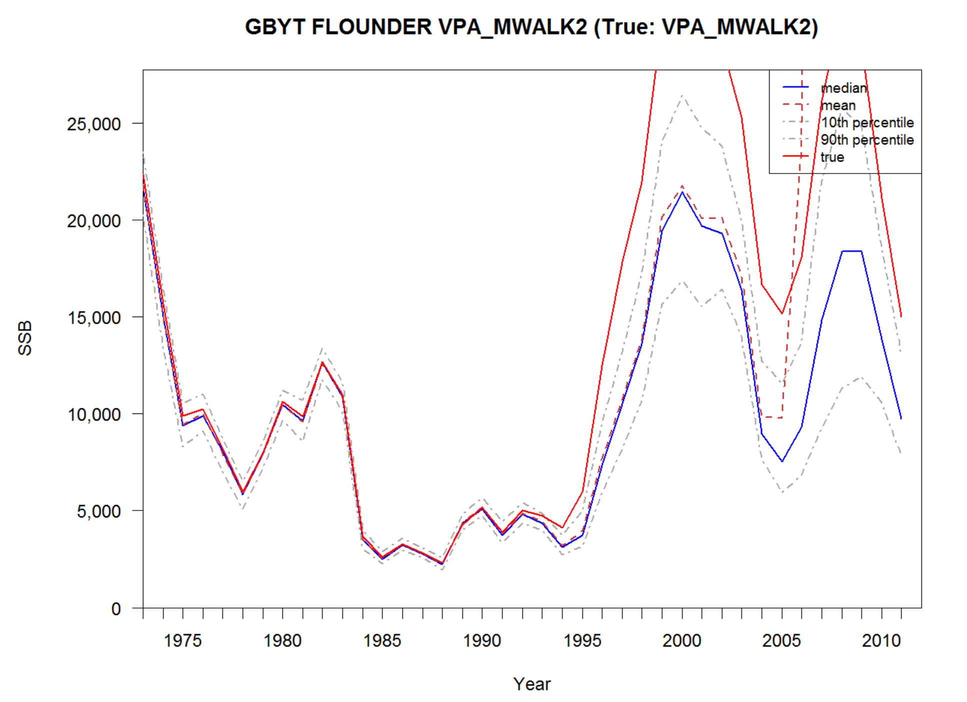
Each type created for both SSB and F

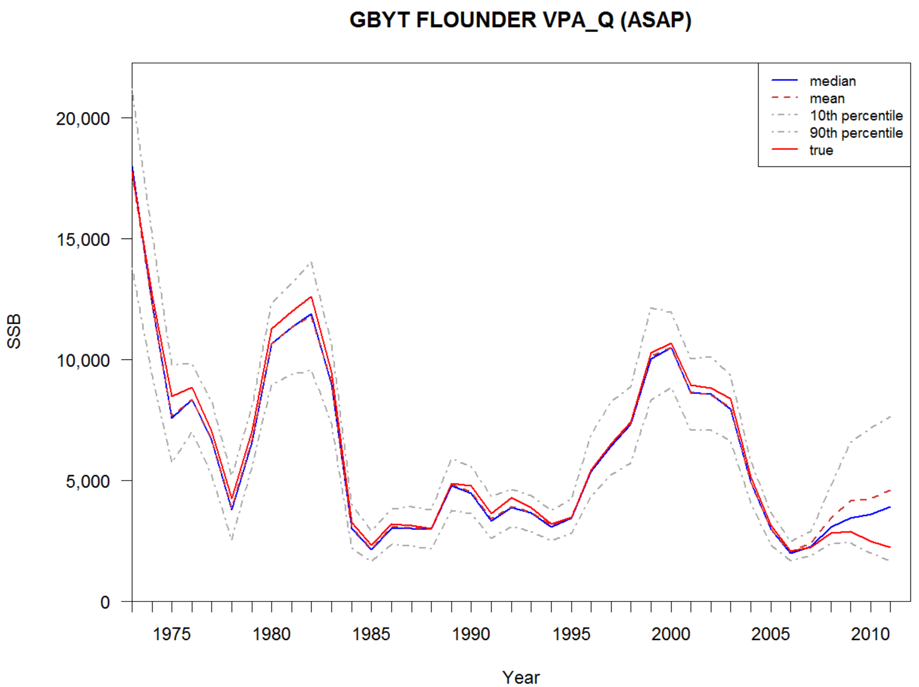
Will show Type 1 for SSB in following slides

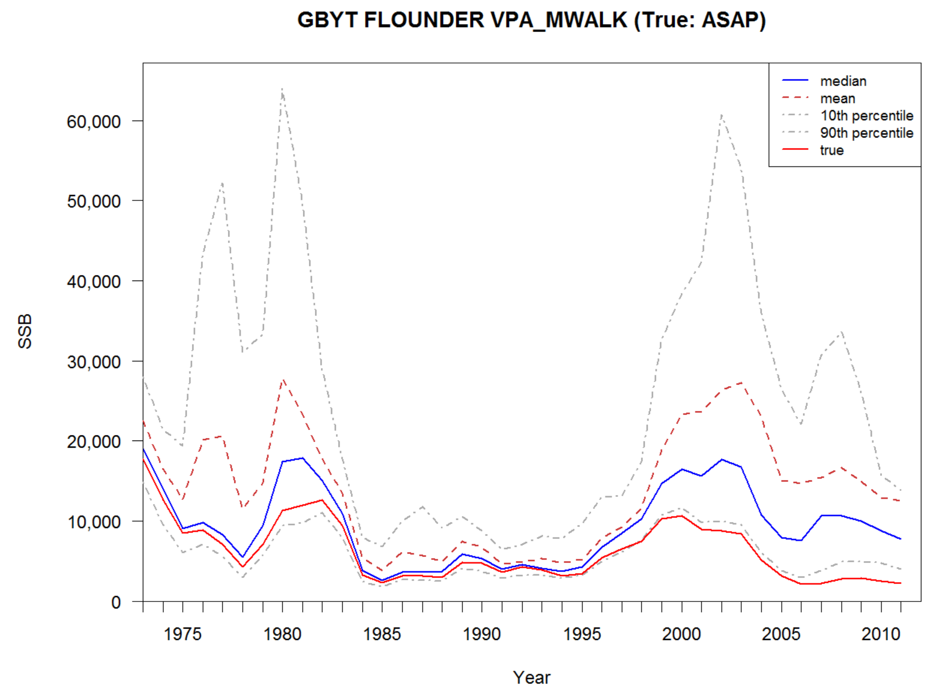




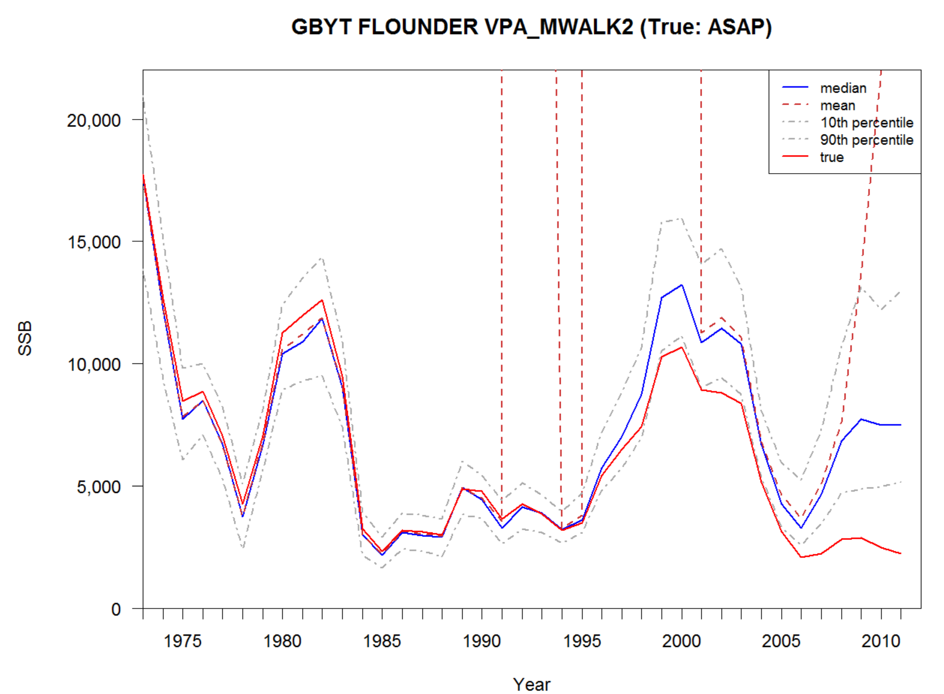
Random walk models against themselves

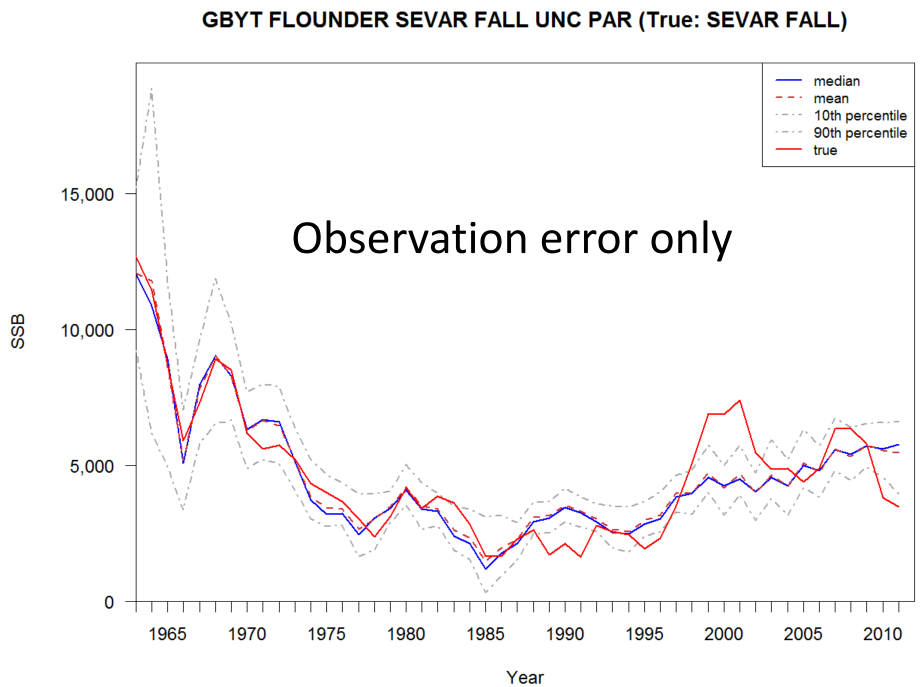


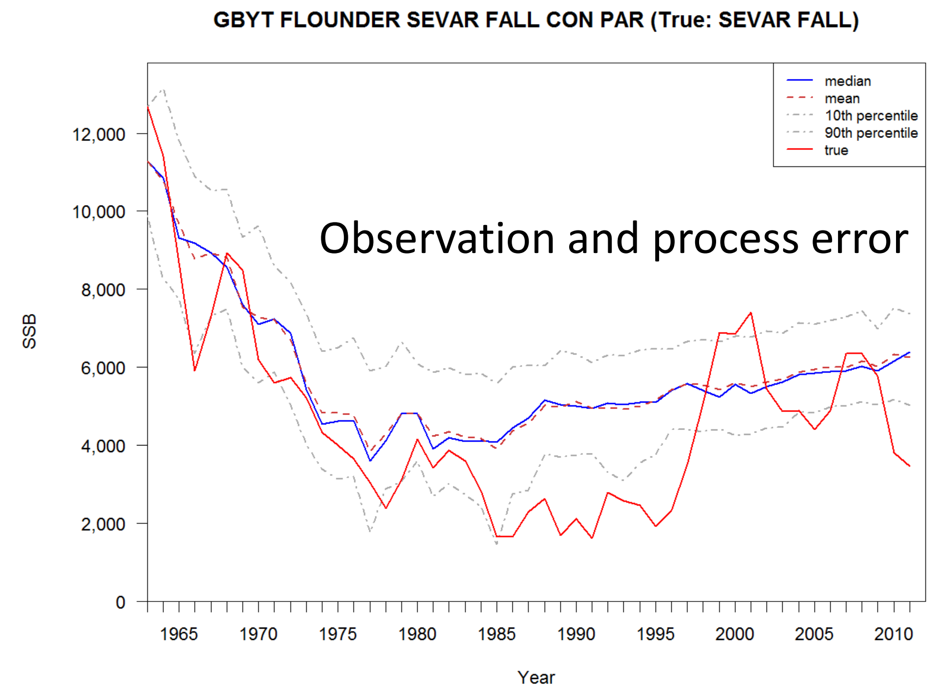




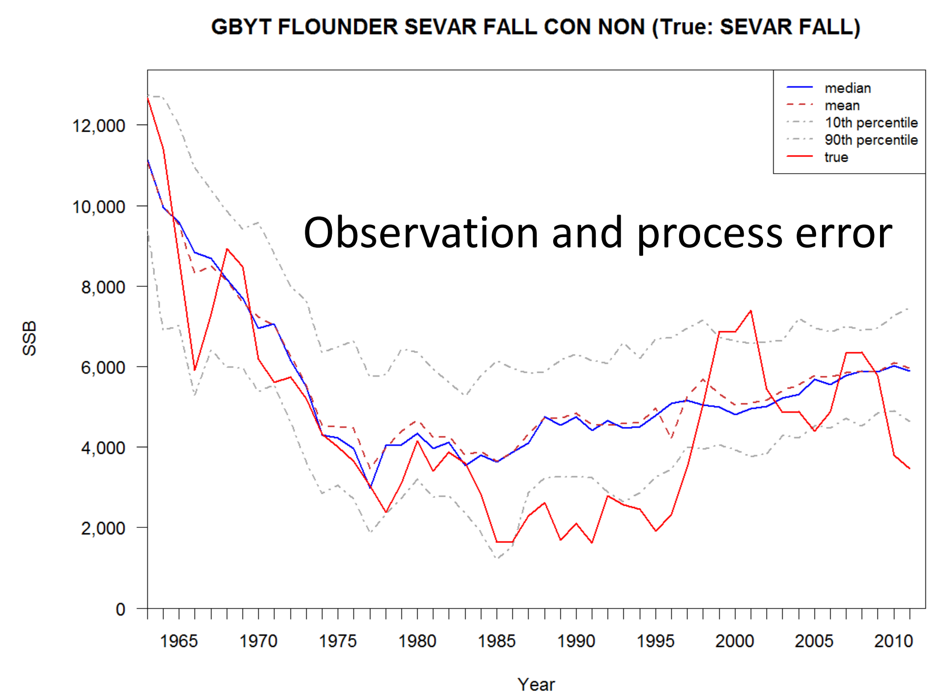
Random walk models against other random walk models

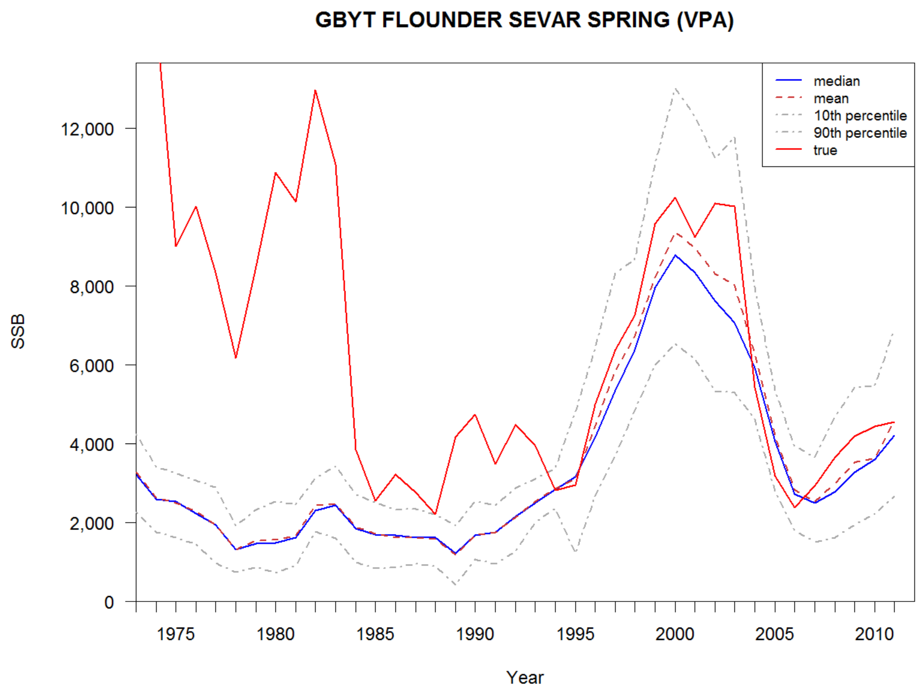


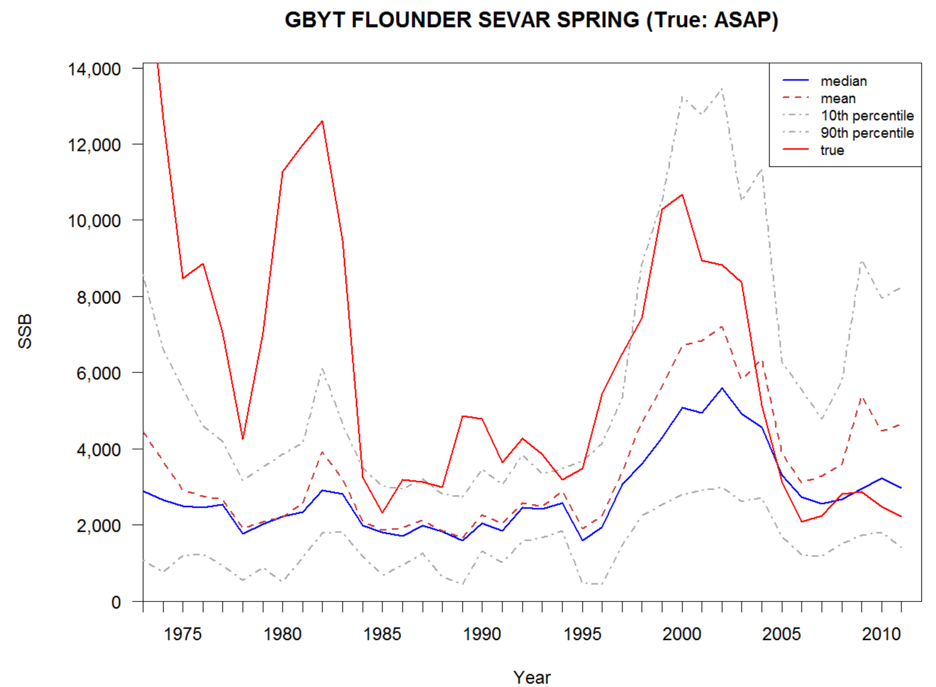




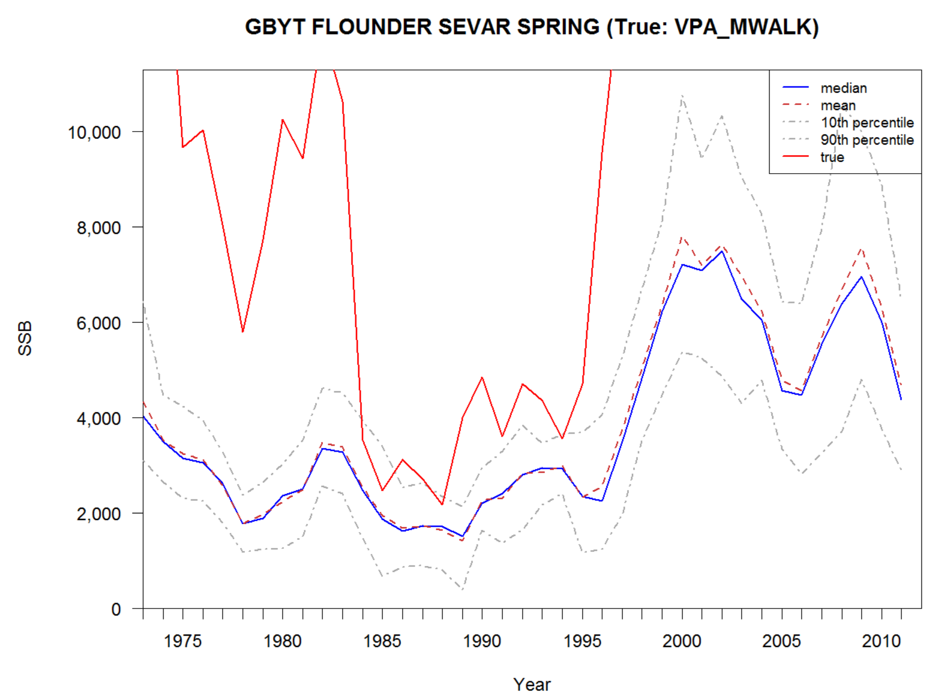
SEVAR against itself
(similar results for Spring)







SEVAR against other models
note scale not directly comparable
(similar results for Fall)



What did we learn?

- Difficult to set up this sort of experiment!

- Random walk models performed poorly against themselves and other random walk models
- SEVAR did well against itself with observation error only, but not as well with process error
- SEVAR performed poorly against random walk or step change

What else did we learn?

- Diagnostics not available
- Catch advice not provided
- Reference points not estimated

- Should these be included in future experiments?

Discussion starters

- How could this exercise be improved?
  - Types of data provided
  - Simulations conducted
  - Metrics compared
  - Logistics
- How can lessons learned be applied in real stock assessments?
  - Should multiple models always/never be used?
  - Guidance for model fitting?
  - Guidance for model development?
  - Should parameters always/never change over time?

Generating simulated stocks for between model comparisons

Tim Earl

Cefas

Introduction

- Process overview
- Model details
- Case studies (North Sea Cod and Herring)
  - Initial model fit
  - Simulation with observation error
- Process error
  - Simulation with observation and process error

Process overview

Purpose:

Compare the robustness of stock assessment models to data generated from other stock dynamics assumptions than their own

Cefas

Models fitted

XSA

(Extended Survivors Analysis)

VPA-derived

F>M
catch assumed exact (no gaps)
requires tuning index

Iterative algorithm terminating when Fs converge between iterations

SAM

(State-space Assessment Model)

Random effects:
lnF follows random walk, correlated across ages
lnN noise term independently normal

q_a for each index
Proc. variance: σ_F, σ_R, σ_S
Obs variance: $\sigma_{C,a}$, $\sigma_{S,a}$
S/R params: α, β
Corr. Param: ρ

Includes catch scaling

Stoch-ASPM (SCA)

(Stochastic age-structure production model)

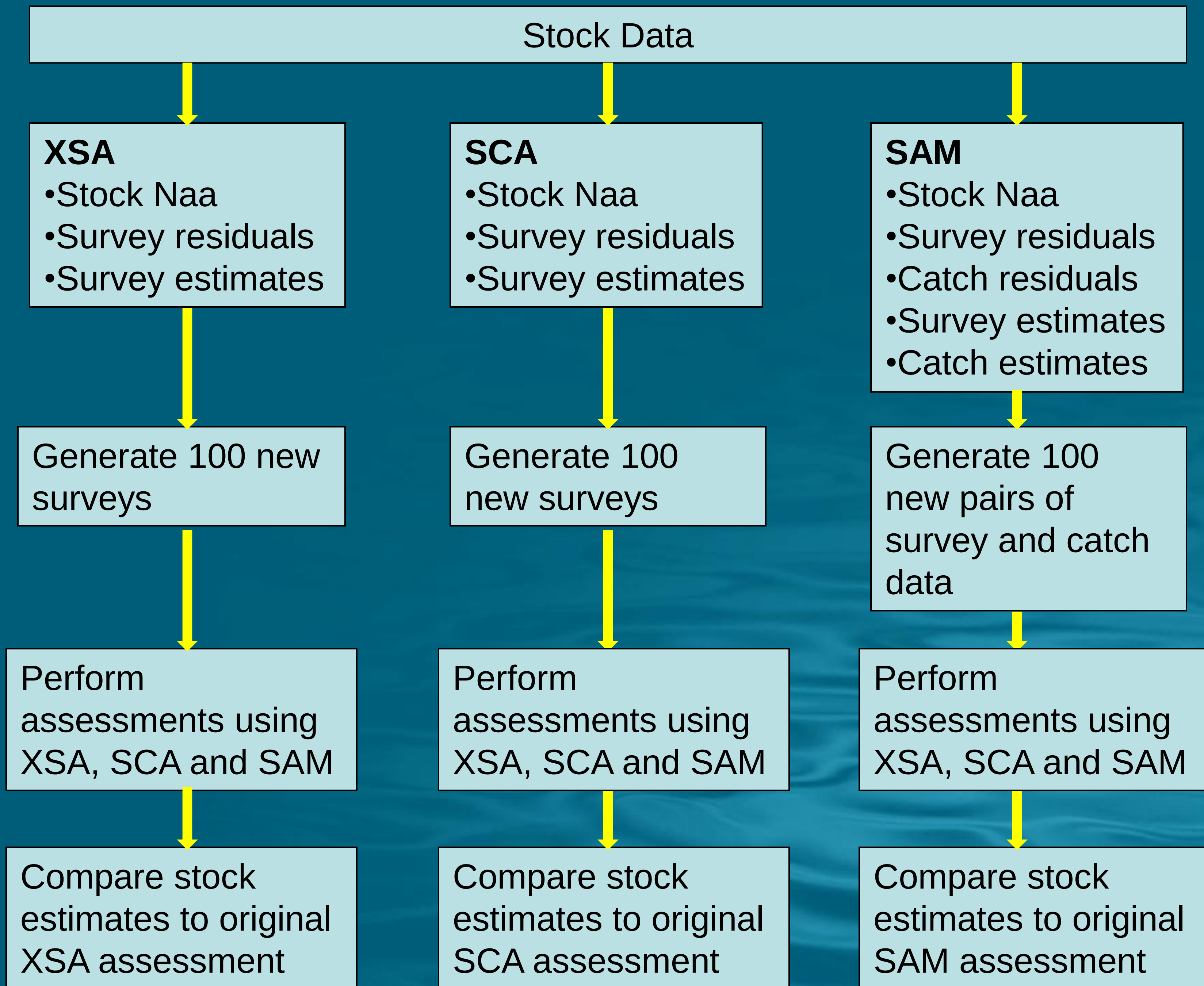
Estimates Fy, Ry-deviates

Selectivity-at-age in block periods, index q constant

S/R included – can be “free”

Penalised likelihood:
logN for total catch and index
Adjusted logN for age-comps

Includes catch scaling


Stock Data
XSA
•Stock Naa
•Survey residuals
•Survey estimates
SCA
•Stock Naa
•Survey residuals
•Survey estimates
SAM
•Stock Naa
•Survey residuals
•Catch residuals
•Survey estimates
•Catch estimates
Generate 100 new surveys
Generate 100 new surveys
Generate 100 new pairs of survey and catch data
Perform assessments using XSA, SCA and SAM
Perform assessments using XSA, SCA and SAM
Perform assessments using XSA, SCA and SAM
Compare stock estimates to original XSA assessment
Compare stock estimates to original SCA assessment
Compare stock estimates to original SAM assessment

Stock Data

XSA
- Stock Naa
- Survey residuals
- Survey estimates

Generate 100 new surveys

- Calculate SD of residuals for each survey and age
- Generate new residuals from normal distribution
- Multiply survey estimates by exp(new residuals)

Perform assessments using XSA, SCA and SAM

Using:
- Simulated survey data
- Actual catch data
- Actual M, Mat, stock weight…

Compare stock estimates to original XSA assessment

Stock Data

SCA
- Stock Naa
- Survey residuals
- Survey estimates

Generate 100 new surveys

- Apply SD of residuals for each survey and age
- Generate new residuals from normal distribution
- Multiply survey estimates by exp(new residuals)

Perform assessments using XSA, SCA and SAM

Using:
- Simulated survey data
- Actual catch data
- Actual M, Mat, stock weight…

Compare stock estimates to original SCA assessment

Stock Data

SAM
- Stock Naa
- Survey residuals
- Catch residuals
- Survey estimates
- Catch estimates

Generate 100 new pairs of survey and catch data

- Apply SD of residuals for each survey and age
- Generate new residuals from normal distribution
- Multiply survey estimates by exp(new residuals)
- Apply SD of residuals for catch at each age
- Generate new residuals from normal distribution
- Multiply catch estimates by exp(new residuals)

Using:
- Simulated survey data
- Simulated catch data
- Actual M, Mat, stock weight…

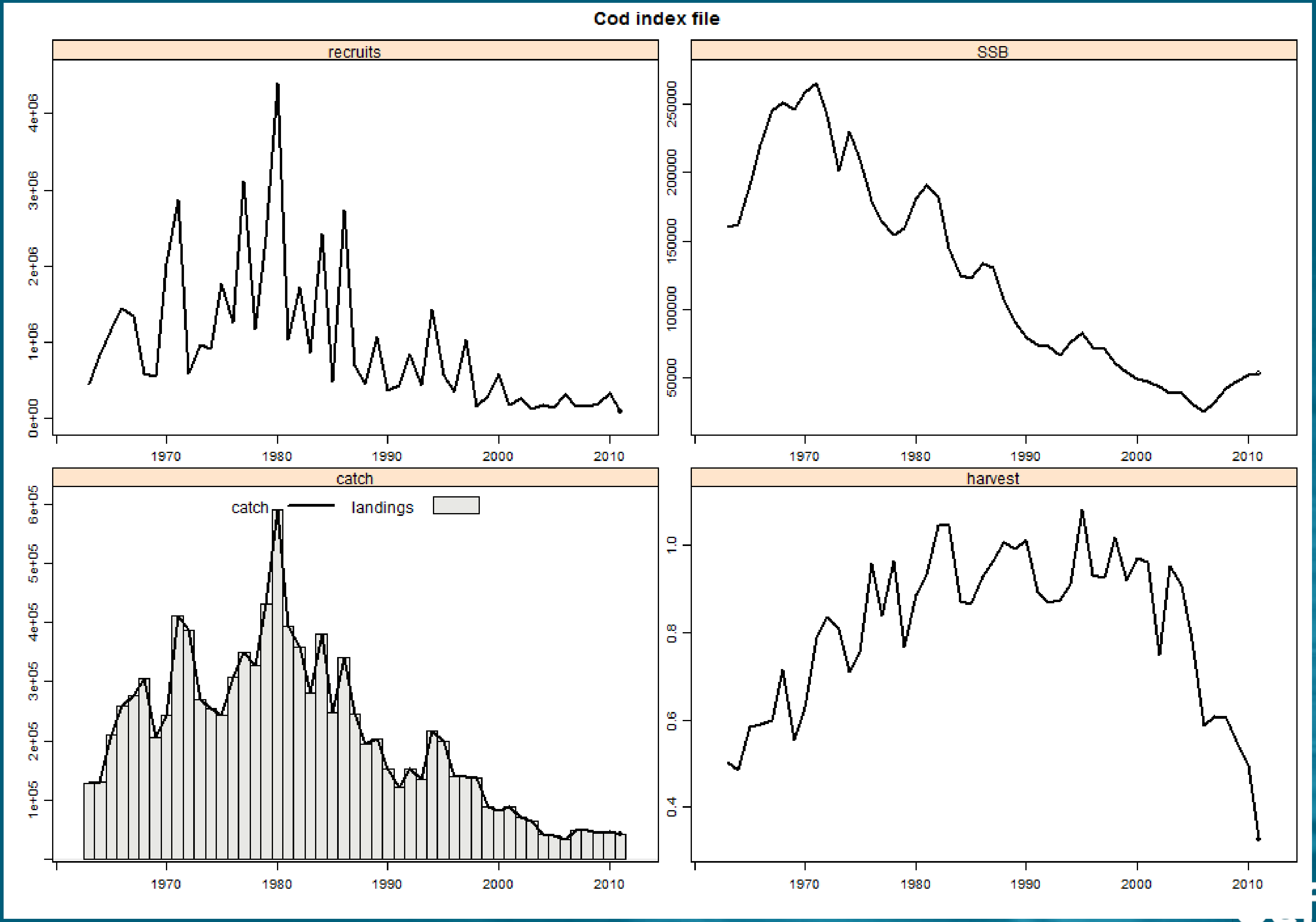
Perform assessments using XSA, SCA and SAM

Compare stock estimates to original SAM assessment

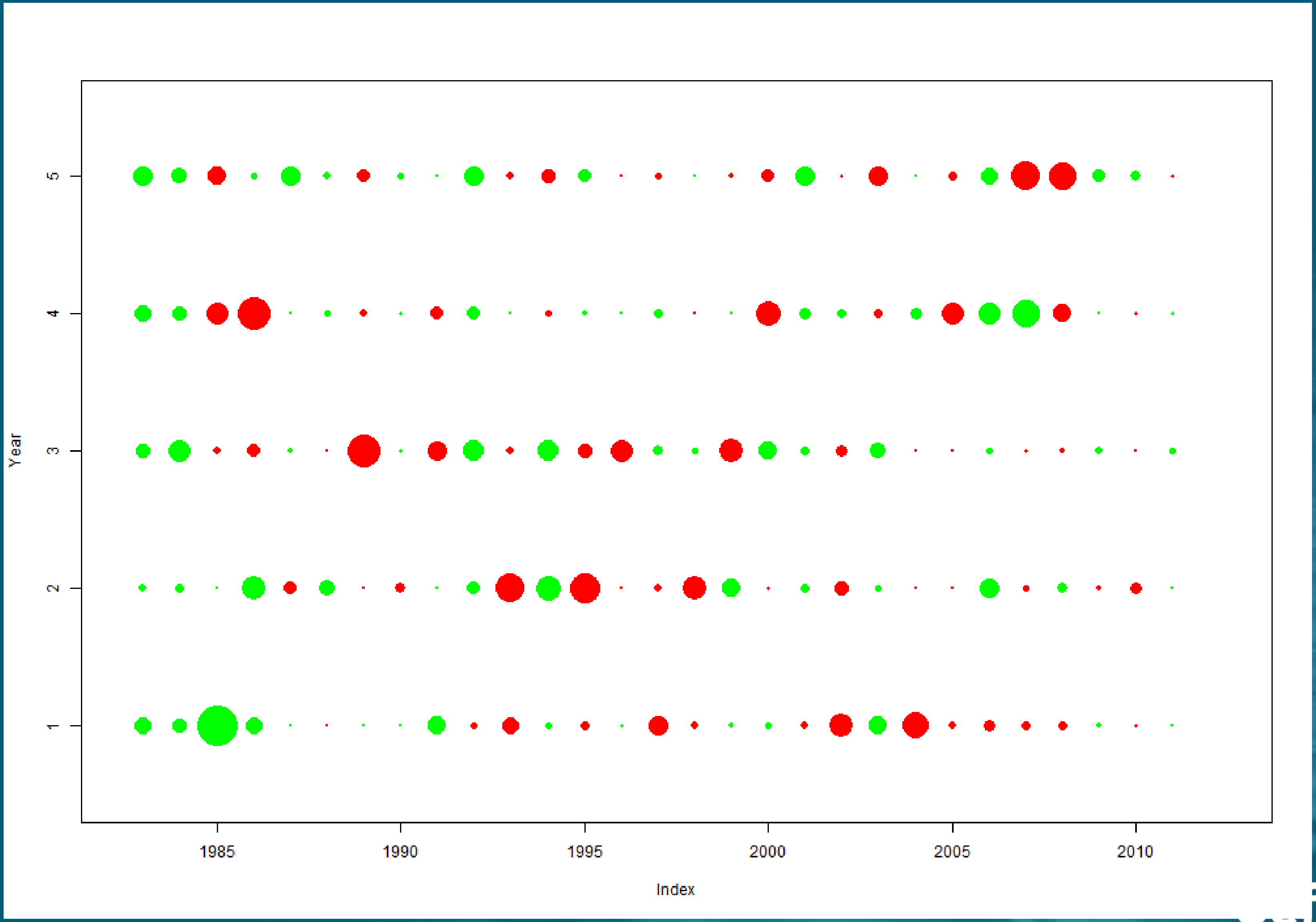
North Sea Cod

- Catch at age, 1963-2011, age 1-7+
- Survey at age, 1983-2012, age 1-5
- Catch multiplier 1993-2005

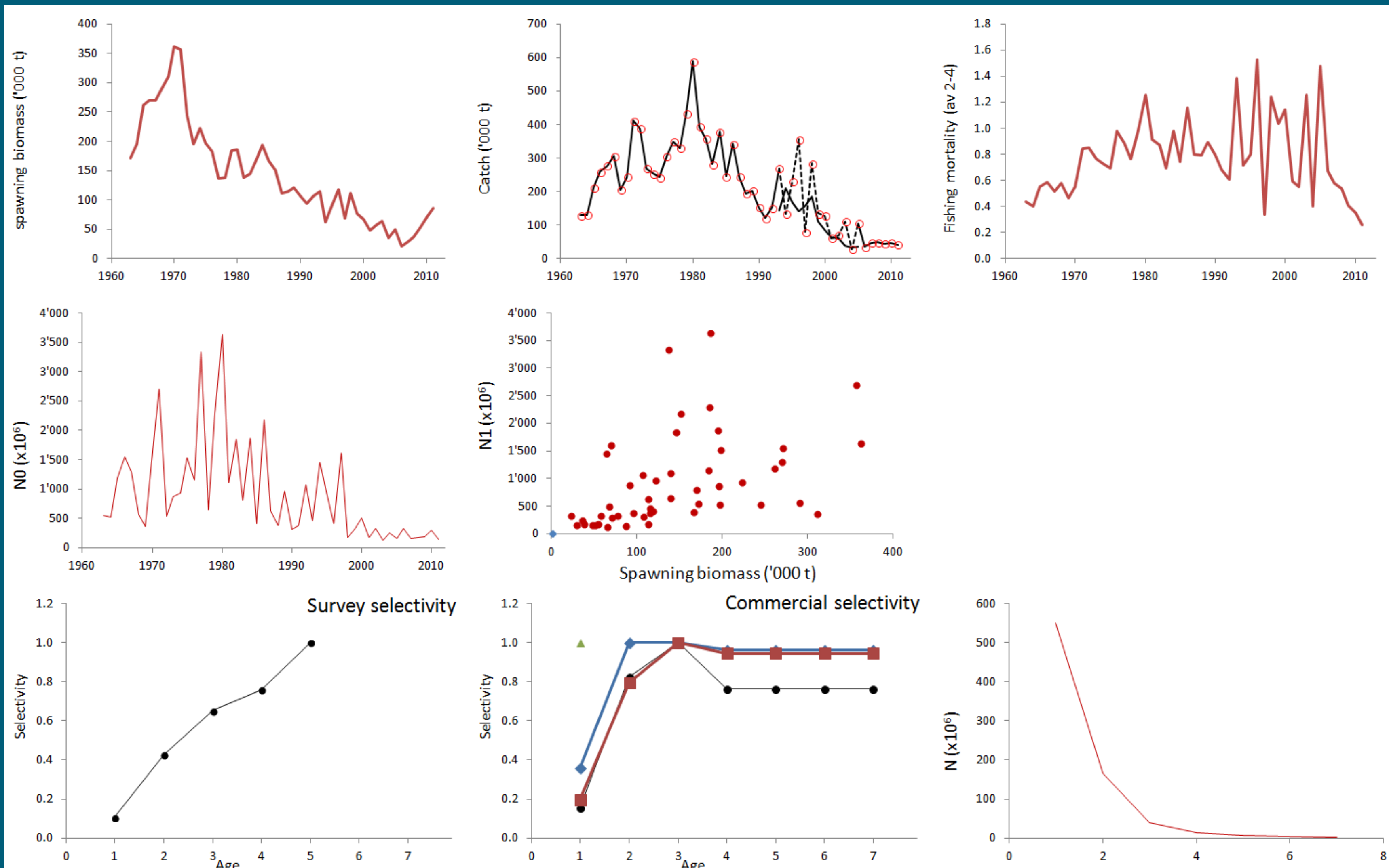
North Sea Cod - XSA



North Sea Cod - XSA



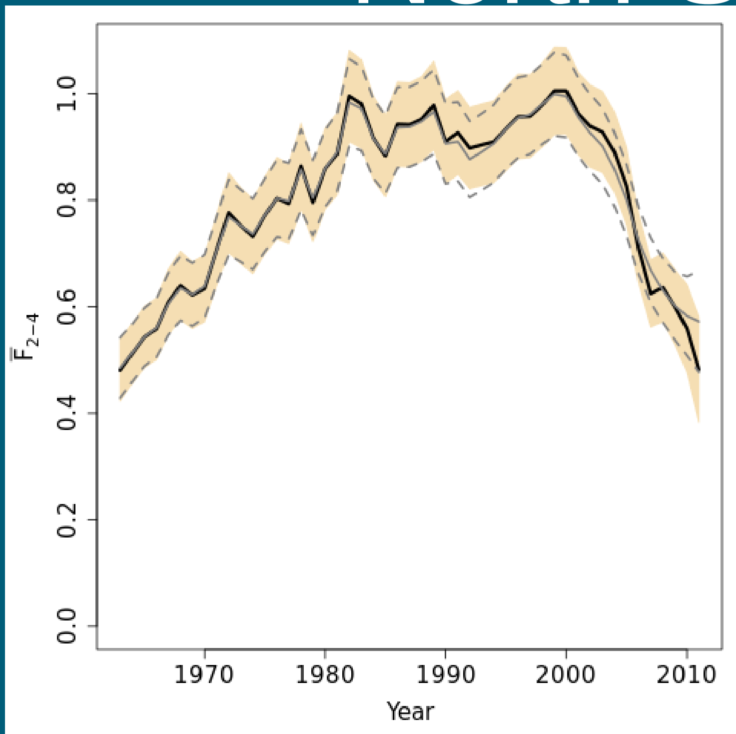
North Sea Cod - SCA

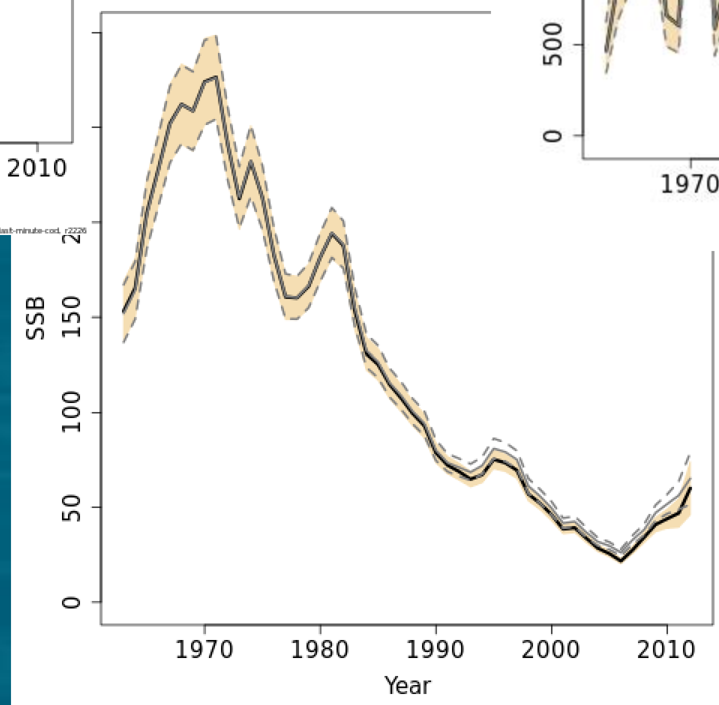


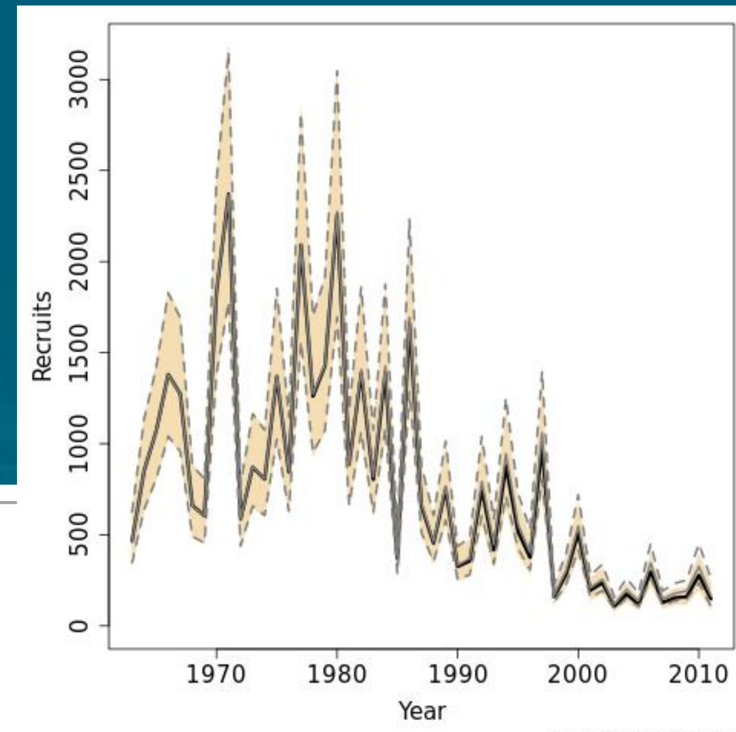
North Sea Cod - SCA

cefas

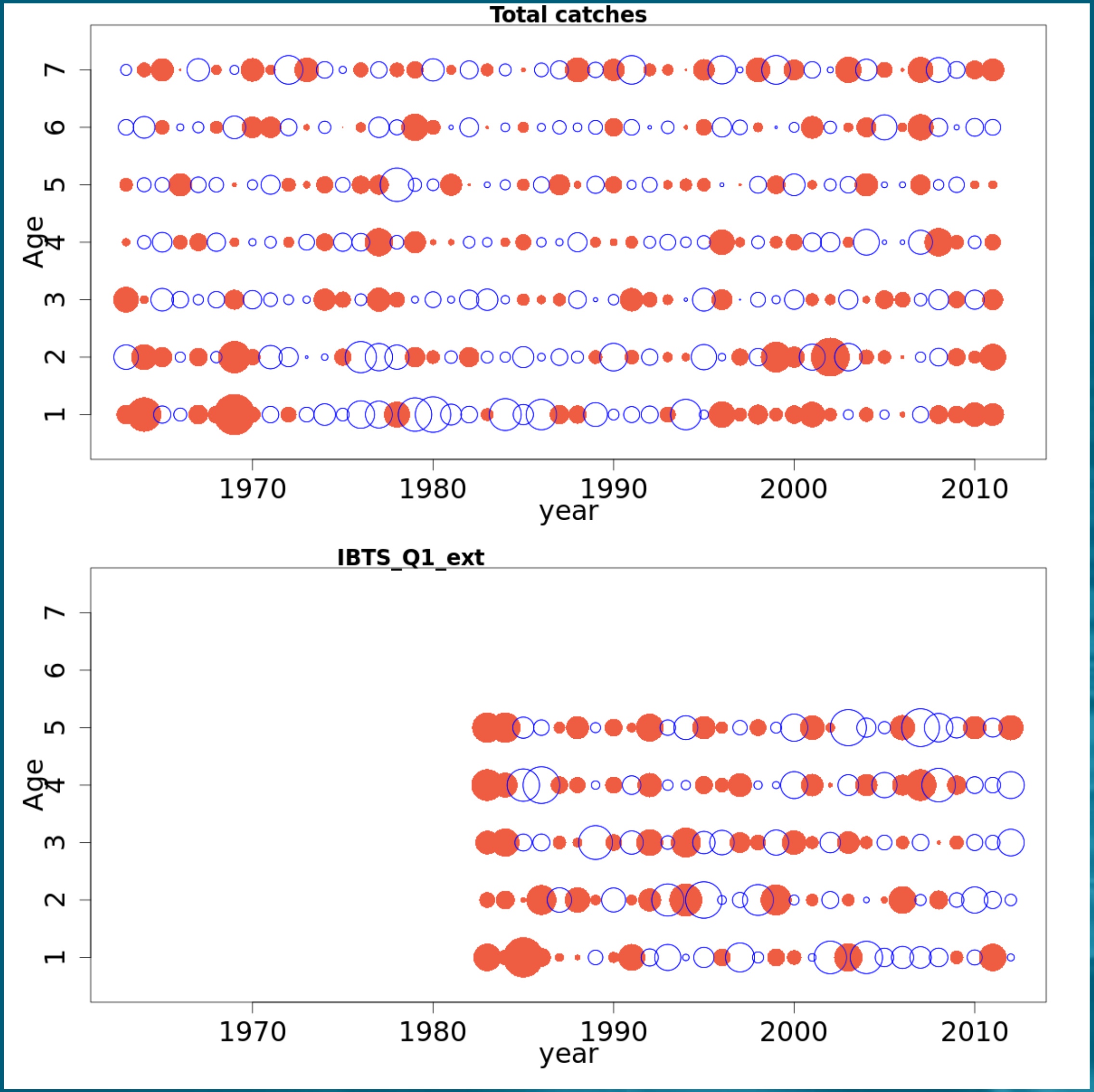
North Sea Cod - SAM







North Sea Cod - SAM



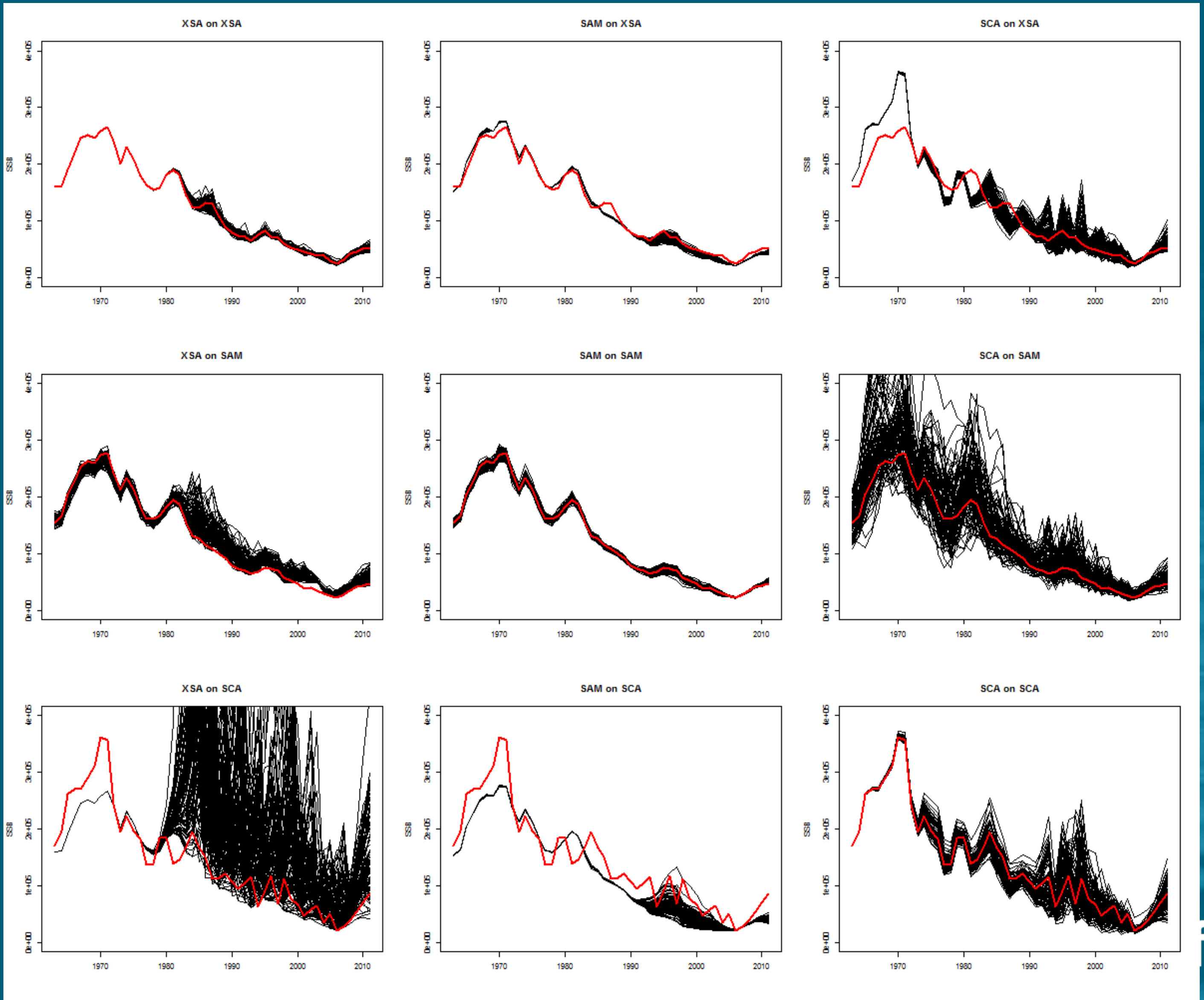
North Sea Cod - SSB

XSA on... SAM on... SCA on...

...data generated from XSA fit

...data generated from SAM fit

...data generated from SCA fit



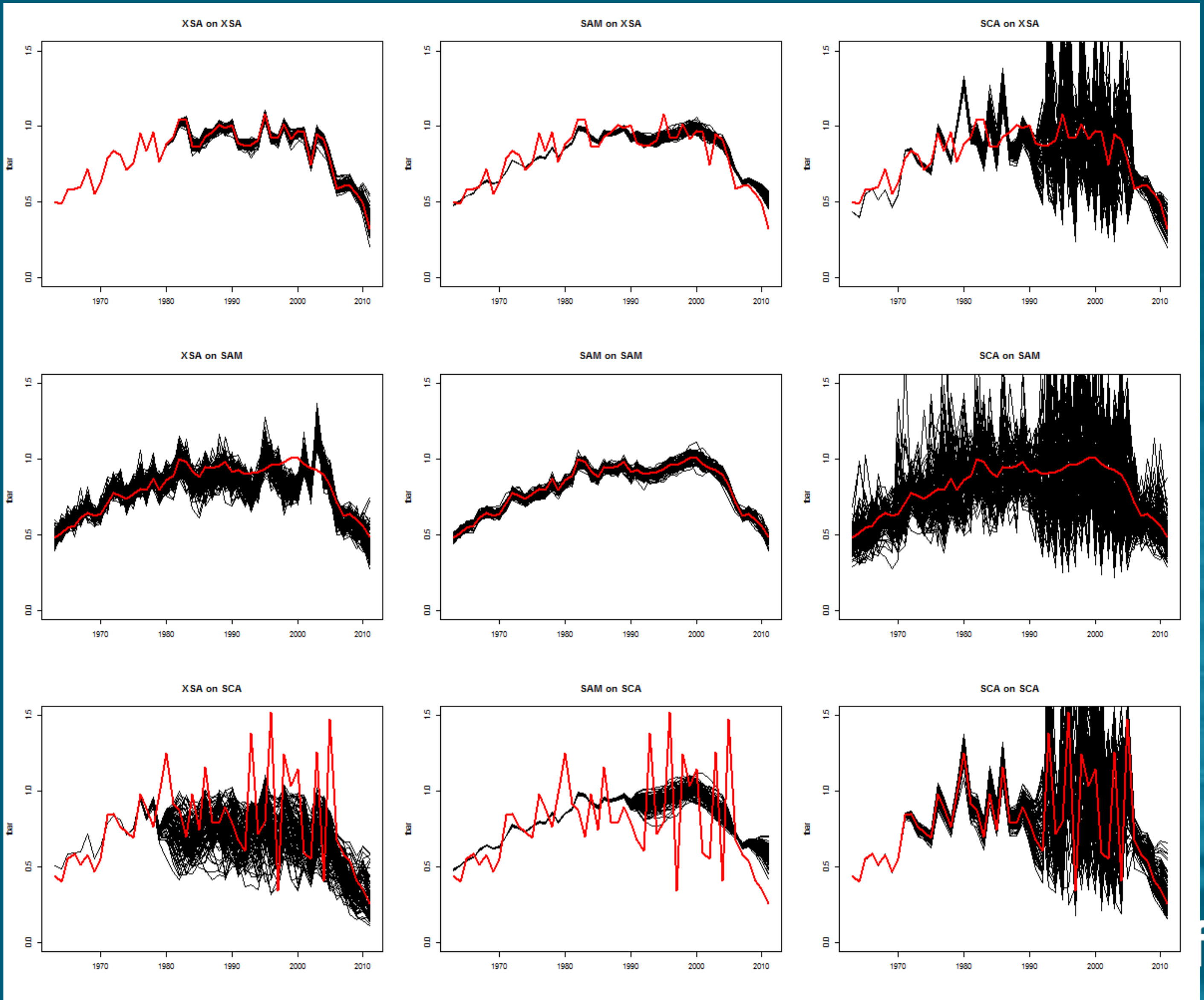
North Sea Cod - Fbar

XSA on... SAM on... SCA on...

...data generated from XSA fit

...data generated from SAM fit

...data generated from SCA fit



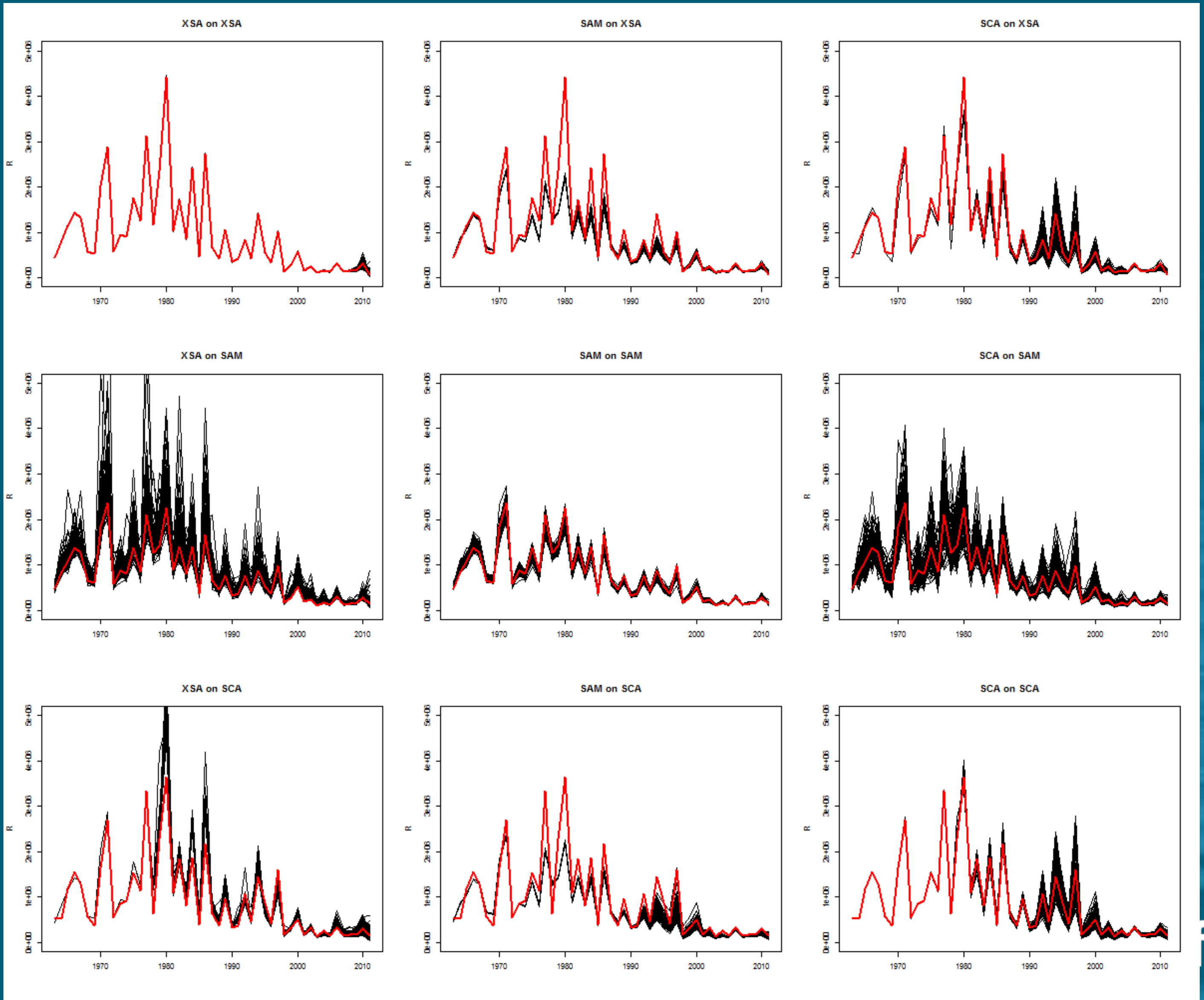
North Sea Cod - Recruitment

XSA on... SAM on... SCA on...

...data generated from XSA fit

...data generated from SAM fit

...data generated from SCA fit

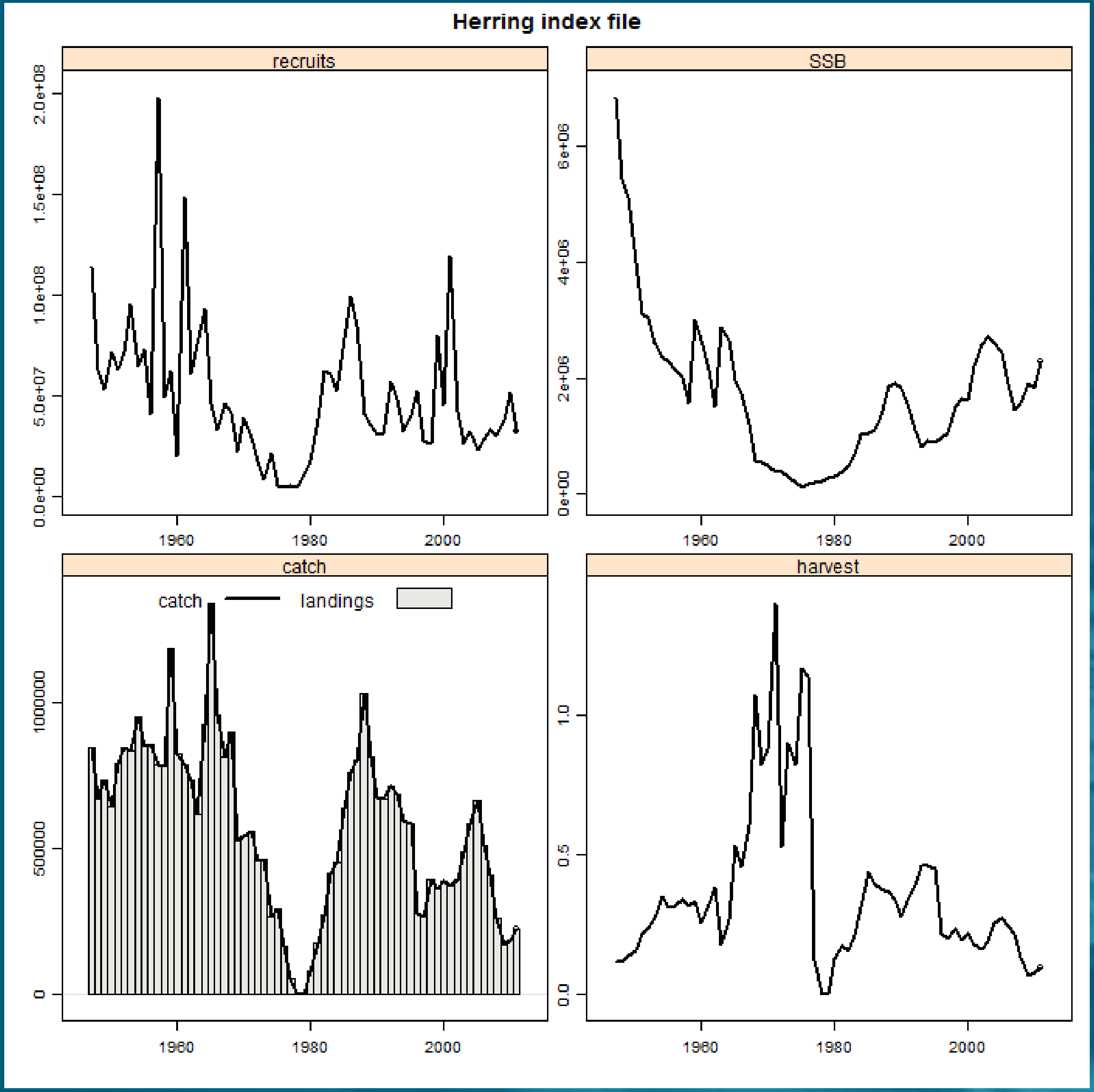


North Sea Herring

- Catch at age, 1947-2011, age 0-8+
  - No catch age composition for ’78 & ’79 (total catch as weight known)
- 3 Surveys at age
  - 1989-2011, age 1-7 (missing age 1 before ’97)
  - 1984-2012, age 1
  - 1992-2012, age 0

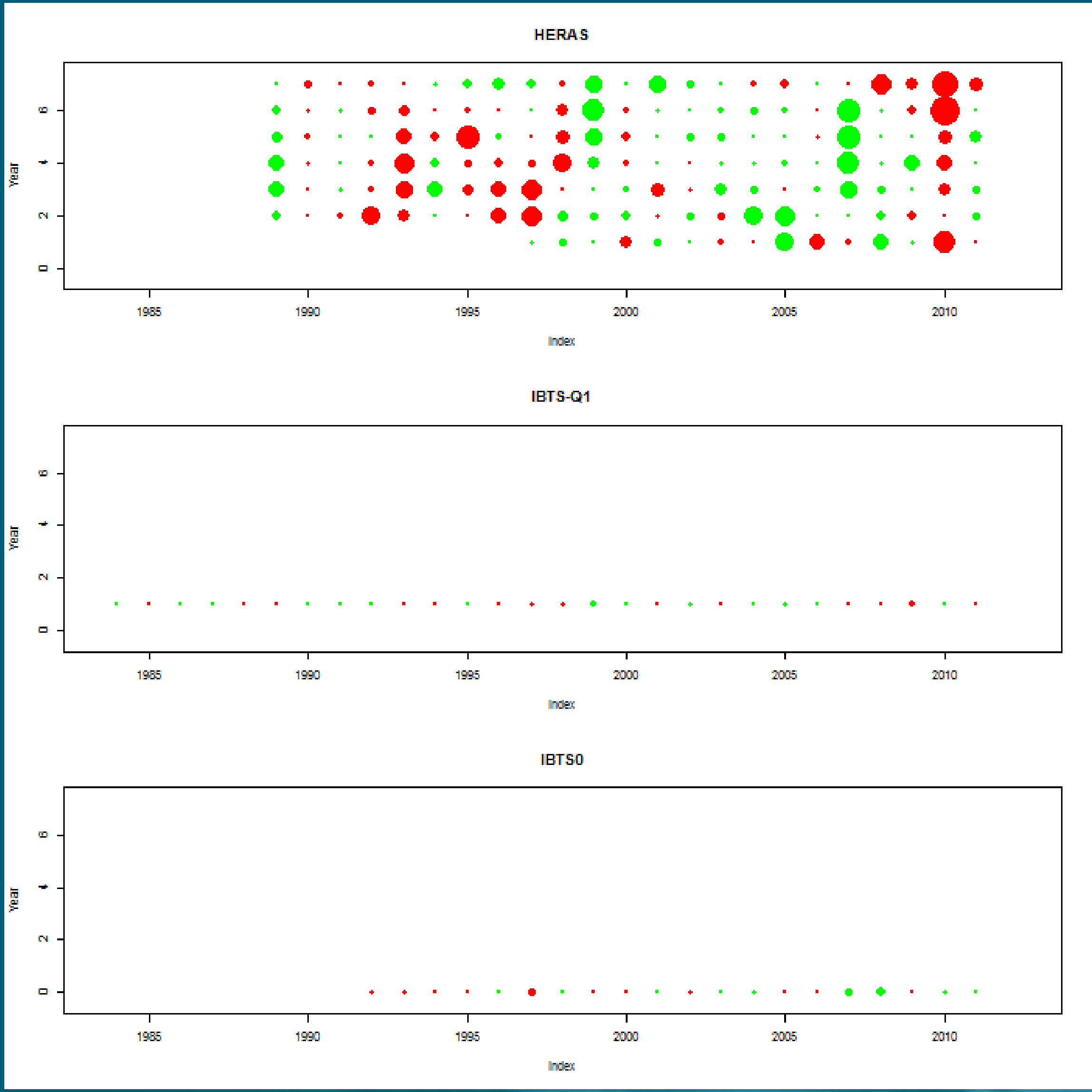
Cefas

North Sea Herring - XSA



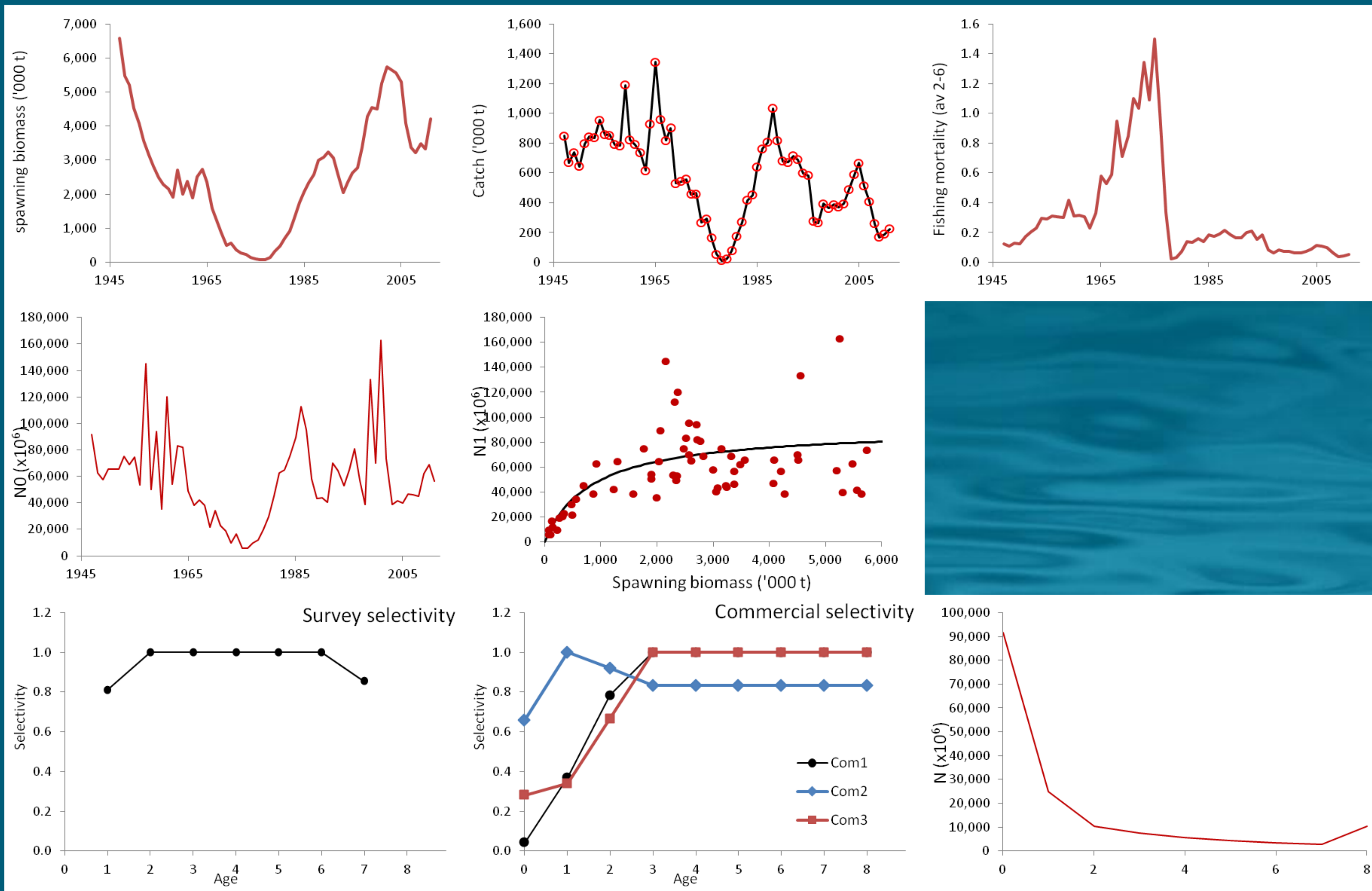
Cefas

North Sea Herring - XSA

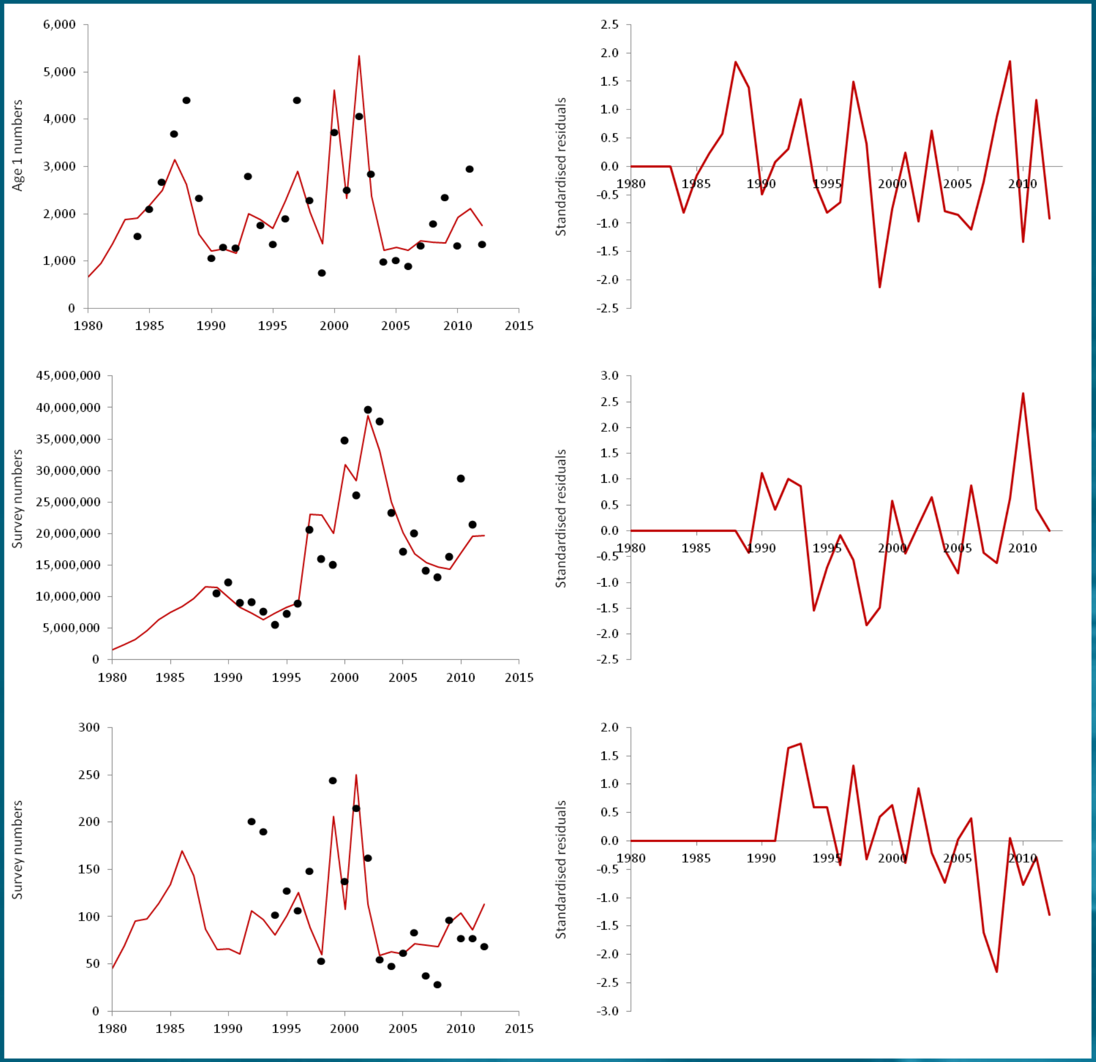


Cefas

North Sea Herring - SCA

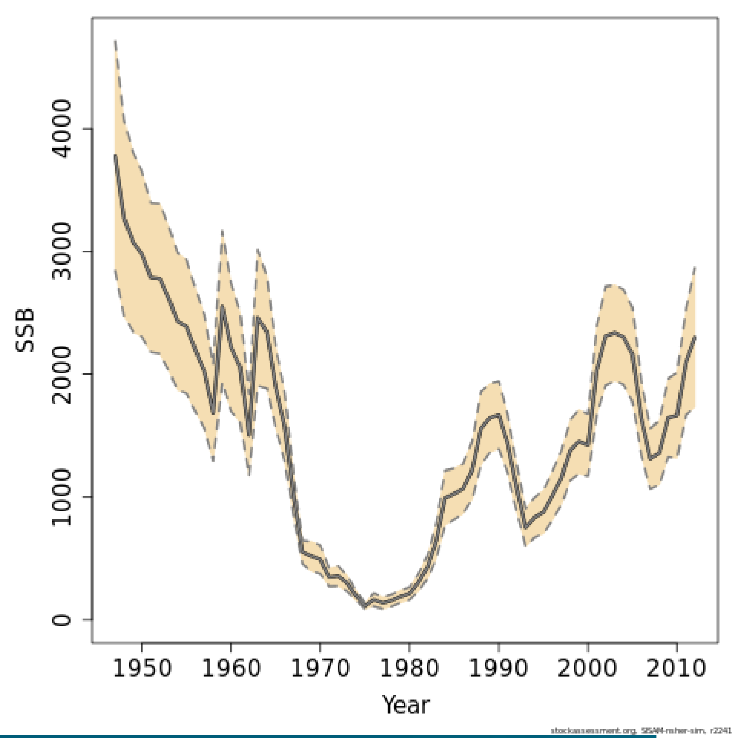


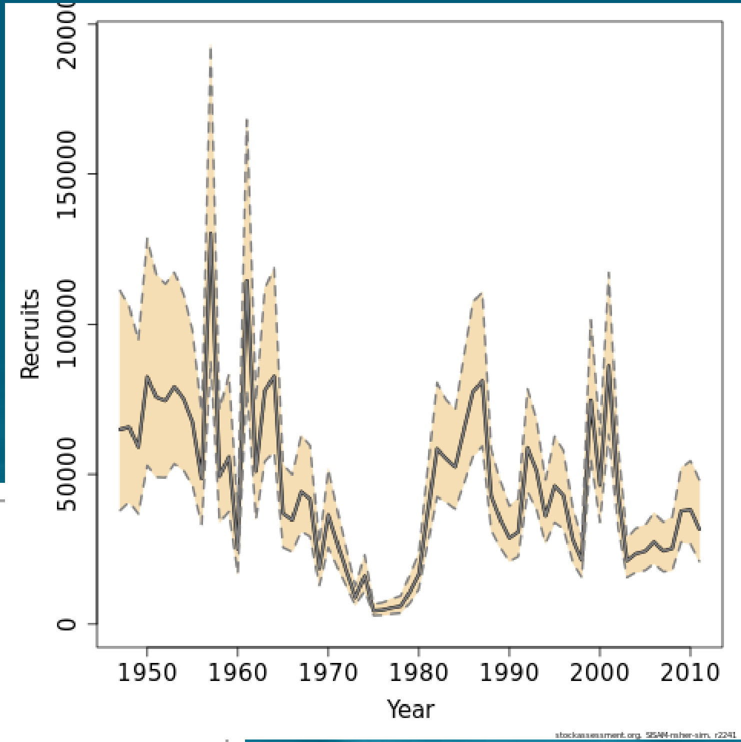
North Sea Herring - SCA

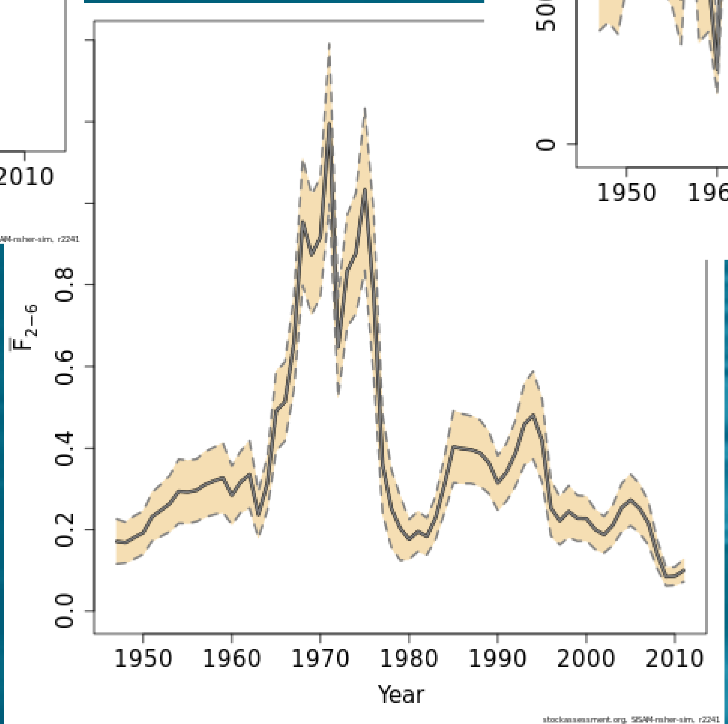


Cefas

North Sea Cod - SAM

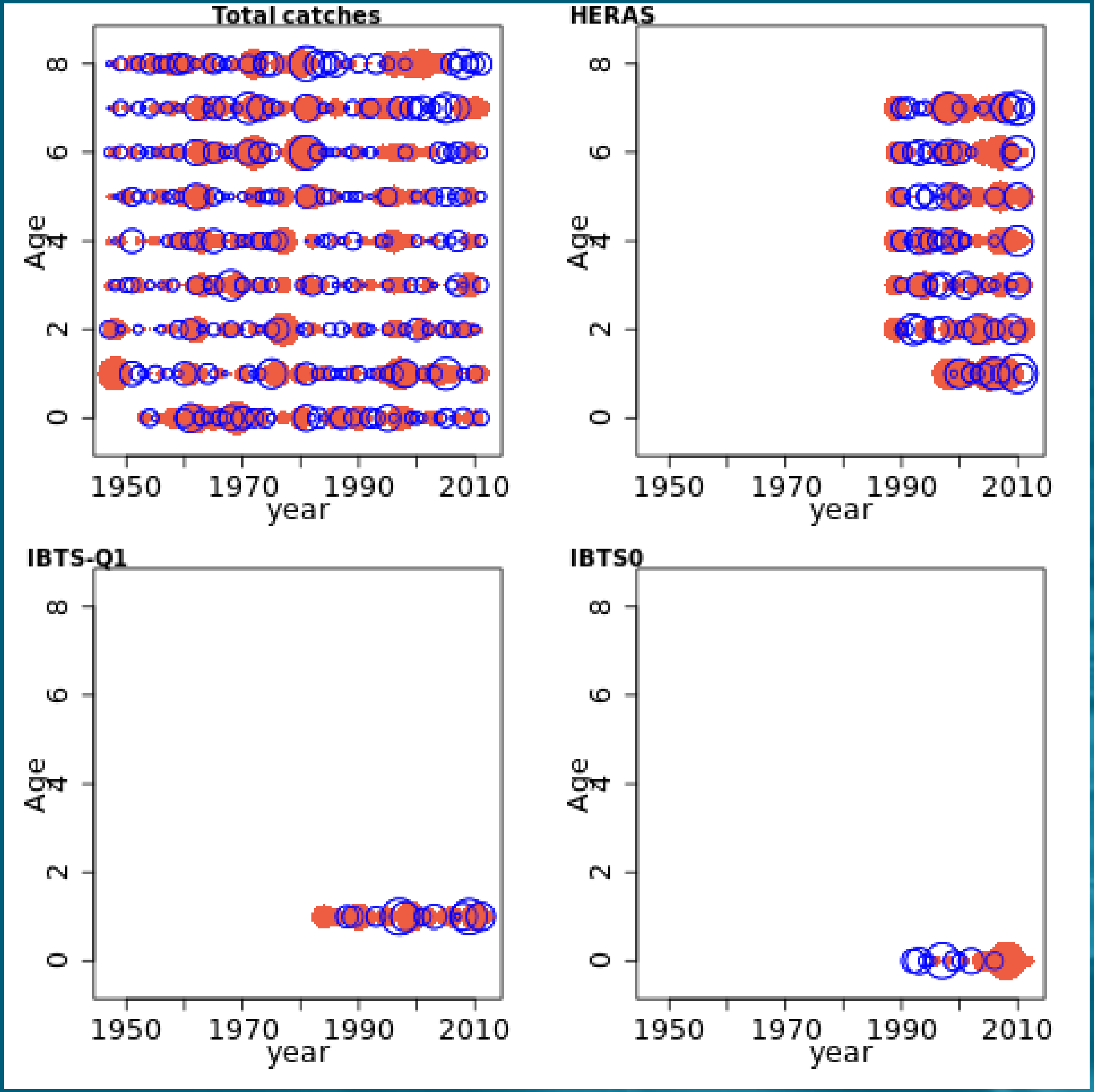






Cefas

North Sea Cod - SAM



Cefas

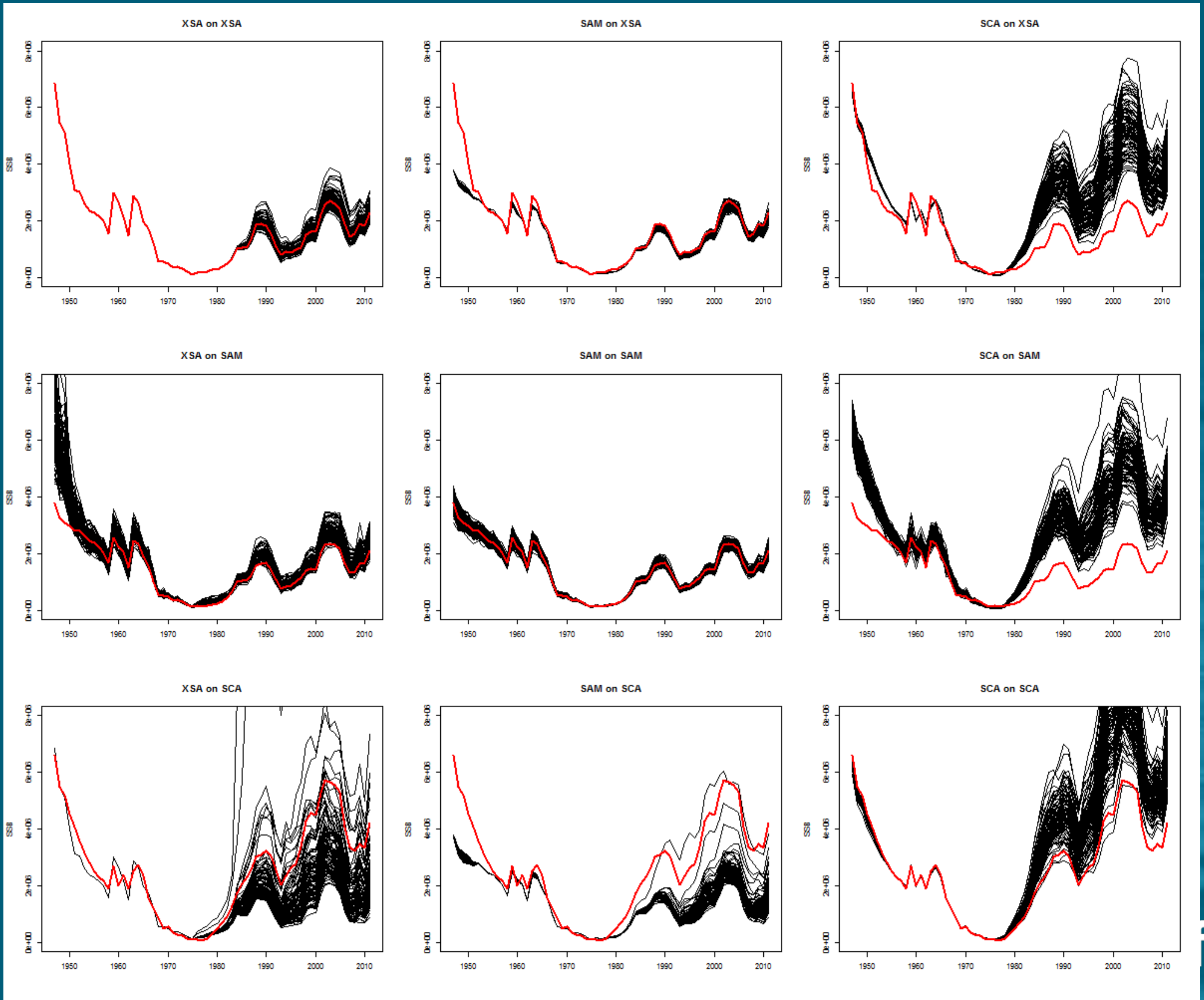
North Sea Herring - SSB

XSA on... SAM on... SCA on...

...data generated from XSA fit

...data generated from SAM fit

...data generated from SCA fit



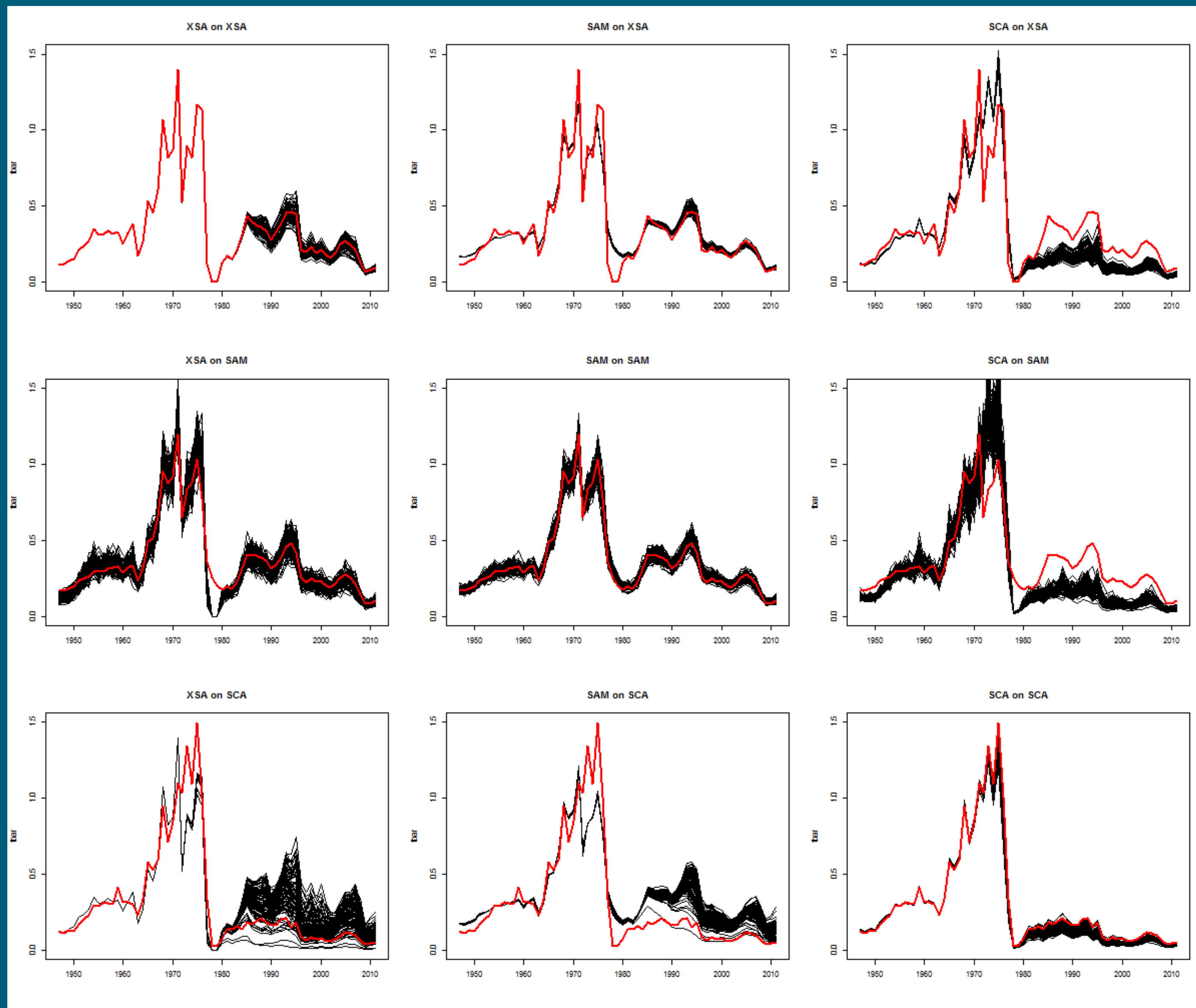
North Sea Herring - Fbar

XSA on... SAM on... SCA on...

...data generated from XSA fit

...data generated from SAM fit

...data generated from SCA fit



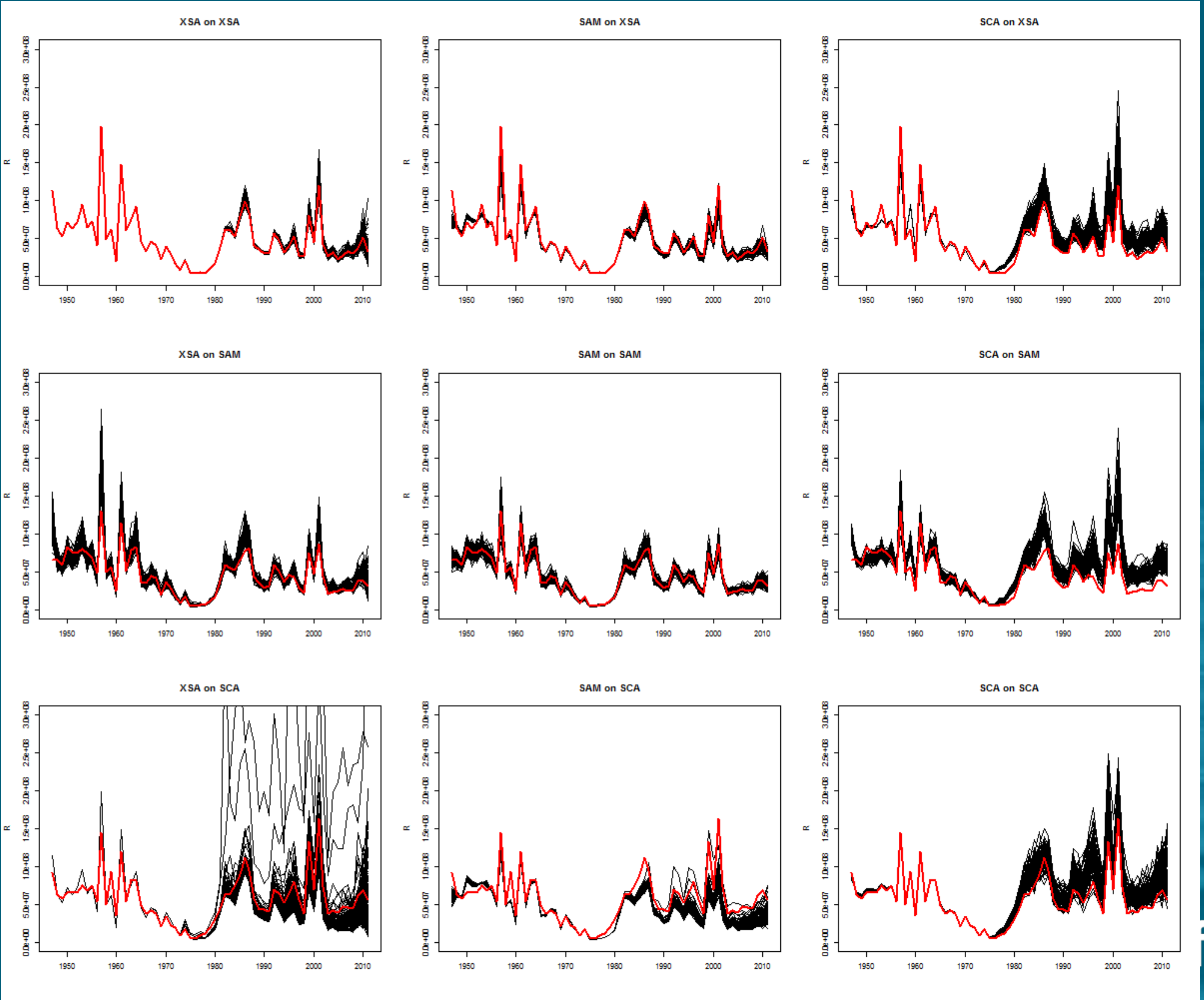
North Sea Herring - Recruitment

XSA on... SAM on... SCA on...

...data generated from XSA fit

...data generated from SAM fit

...data generated from SCA fit



Process Error

Stock Data

SAM
- Stock Naa
- Survey residuals
- Catch residuals
- Survey estimates
- Catch estimates

Generate 100 new pairs of survey and catch data

Perform assessments using XSA, SCA and SAM

Compare stock estimates to resampled parameters

- Use ADMB estimates of co-variance between random effects which defines a multivariate distribution of N and F
- Simulate N_i and F_i from above
- For each set of N_i and F_i catch and survey catch indices are calculated.
- Observation error is added as in the observation error only case

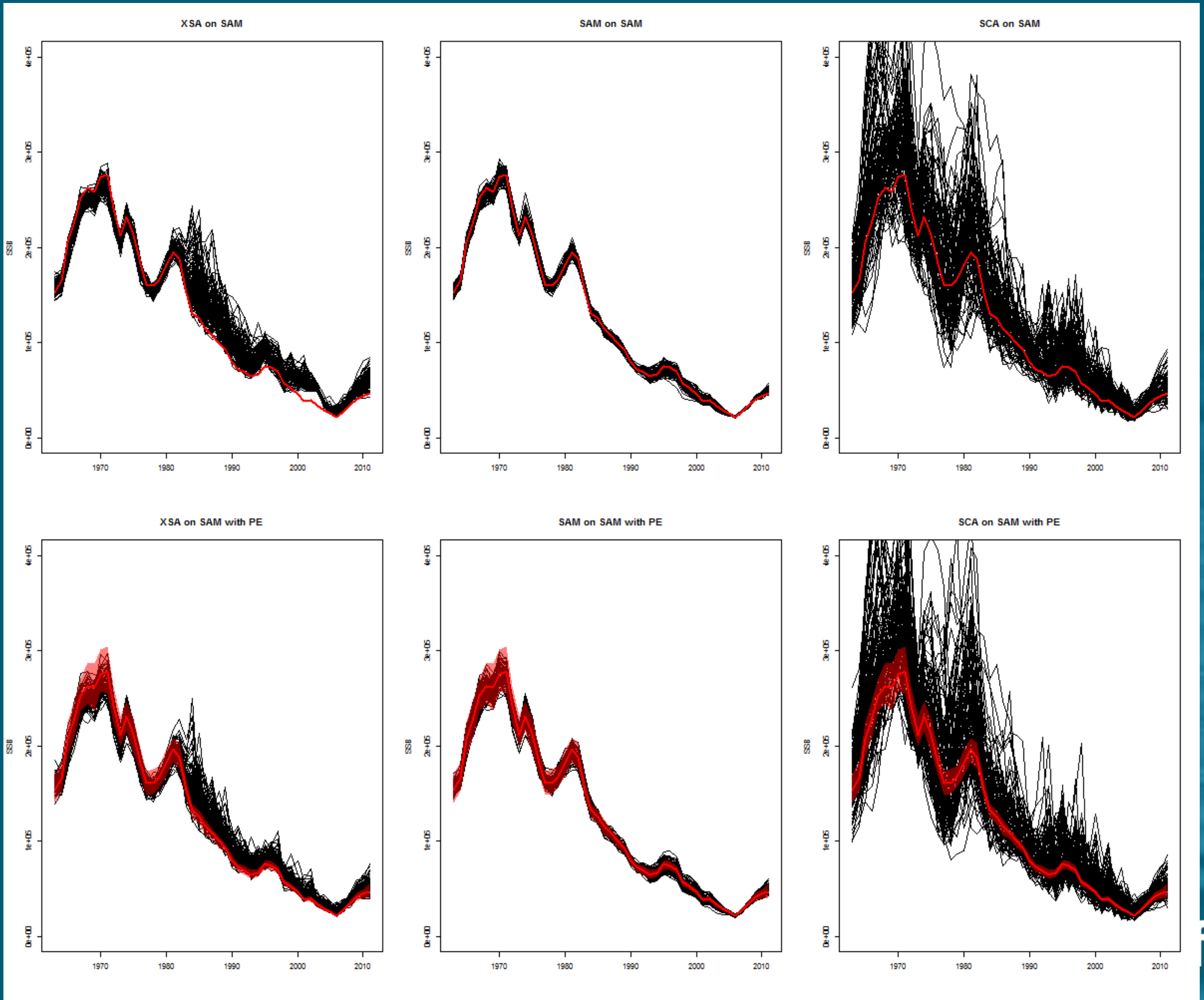
Using:
- Simulated survey data
- Simulated catch data
- Actual M, Mat, stock weight…

North Sea Cod - SSB

XSA on... SAM on... SCA on...

...data generated from SAM fit with obs. error

...data generated from SAM fit with obs. And process error



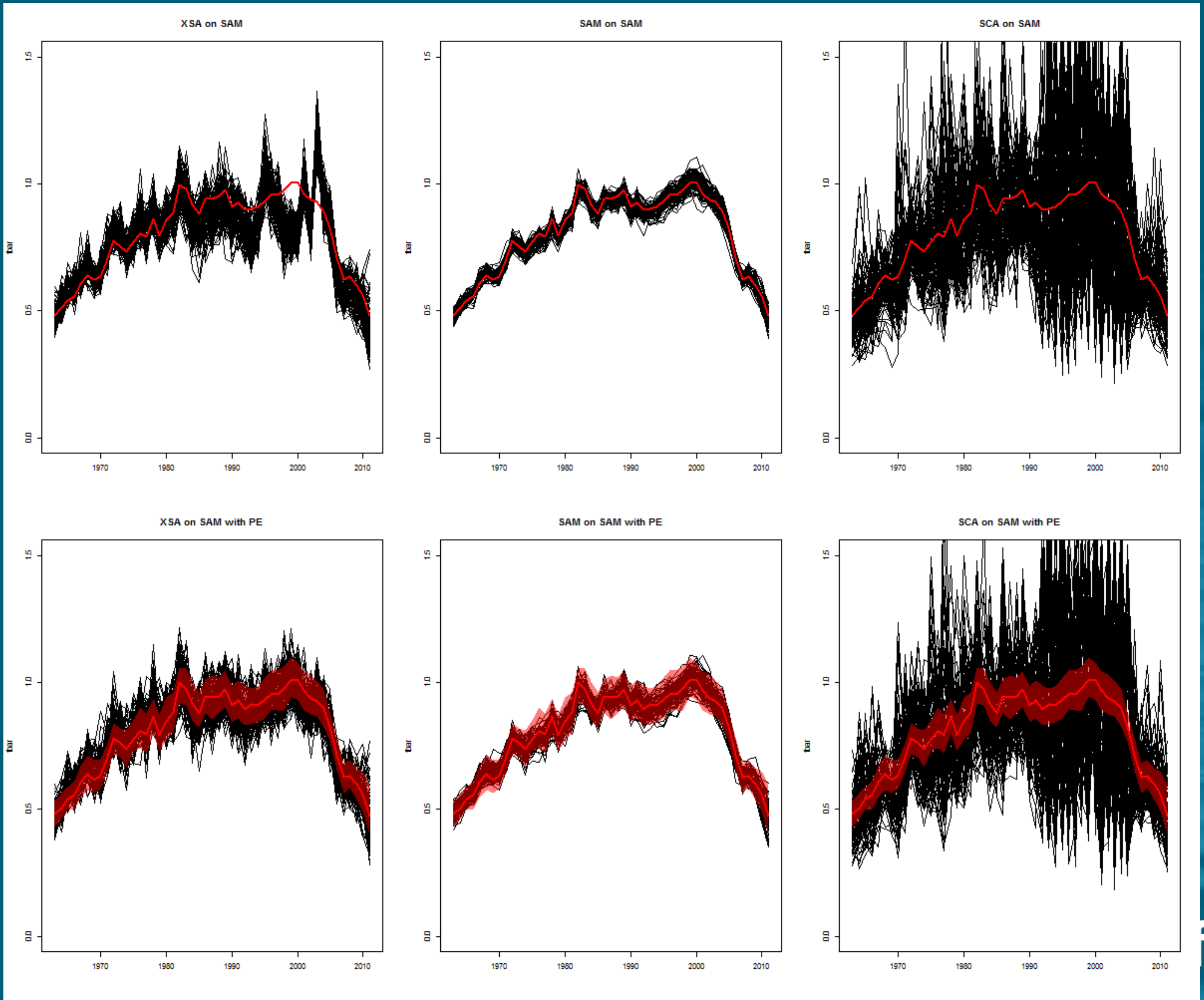
fas

North Sea Cod - Fbar

XSA on... | SAM on... | SCA on...

...data generated from SAM fit with obs. error

...data generated from SAM fit with obs. And process error



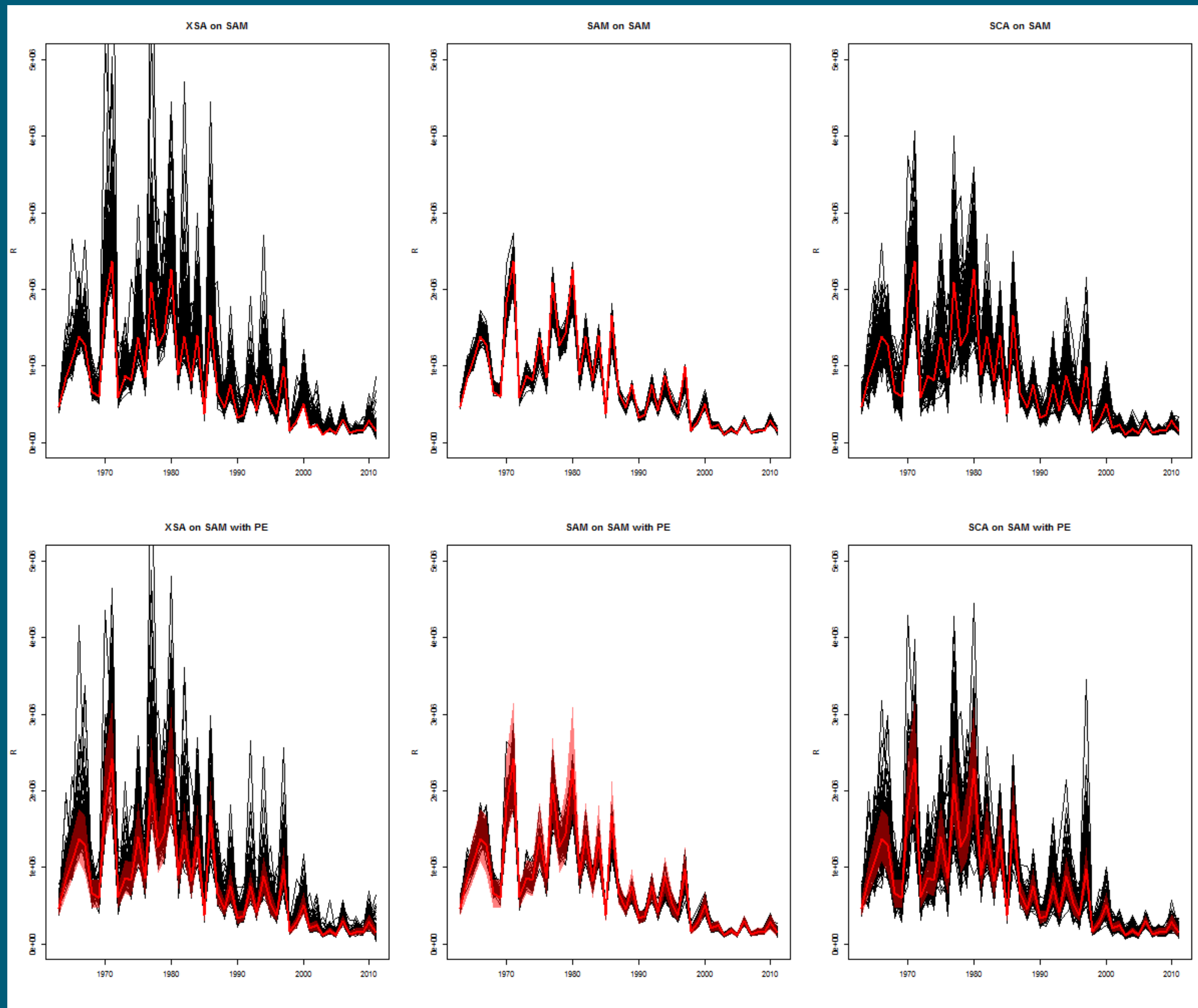
fas

North Sea Cod - Recruitment

XSA on... SAM on... SCA on...

...data generated from SAM fit with obs. error

...data generated from SAM fit with obs. And process error

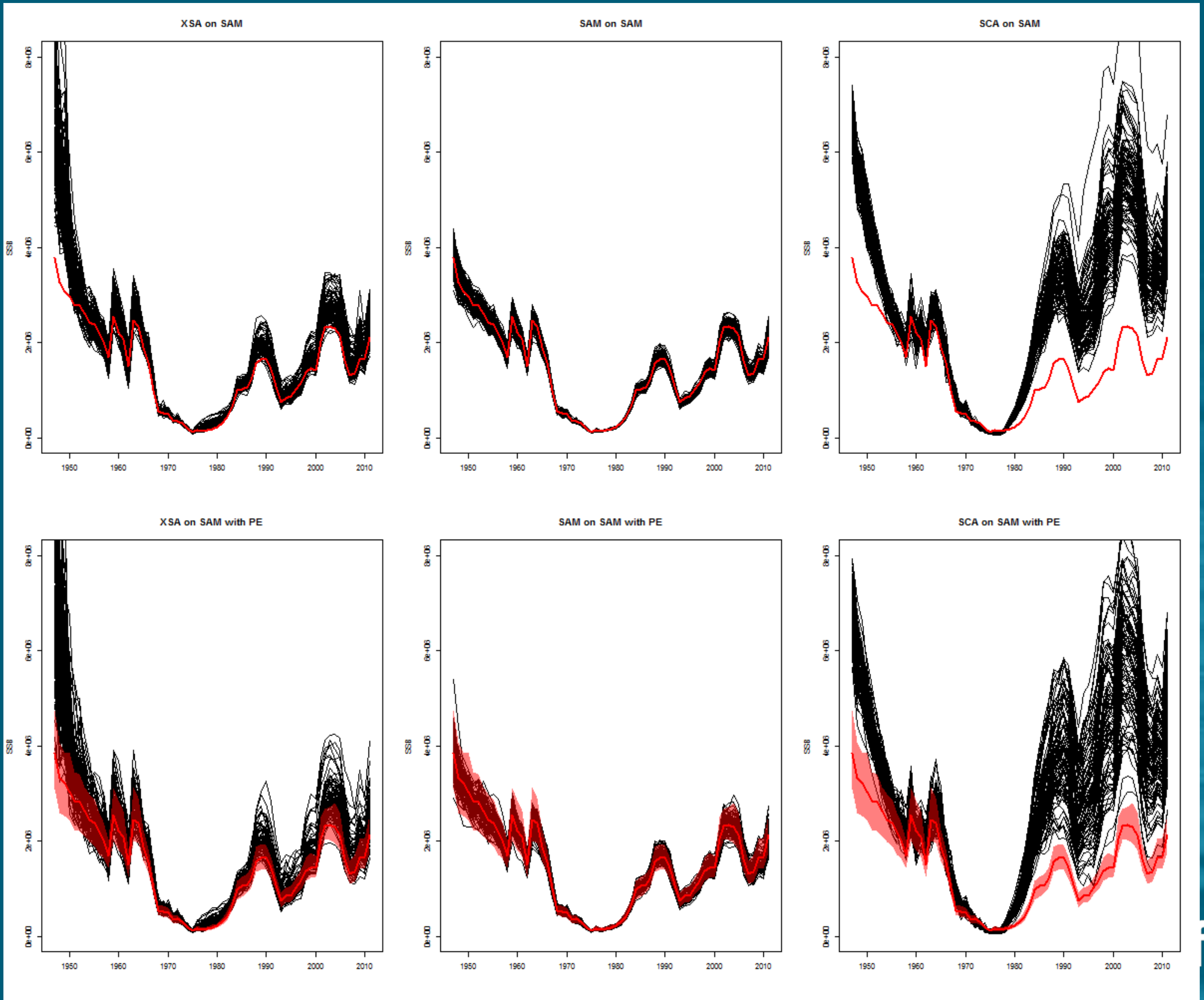


North Sea Herring - SSB

XSA on... SAM on... SCA on...

...data generated from SAM fit with obs. error

...data generated from SAM fit with obs. And process error



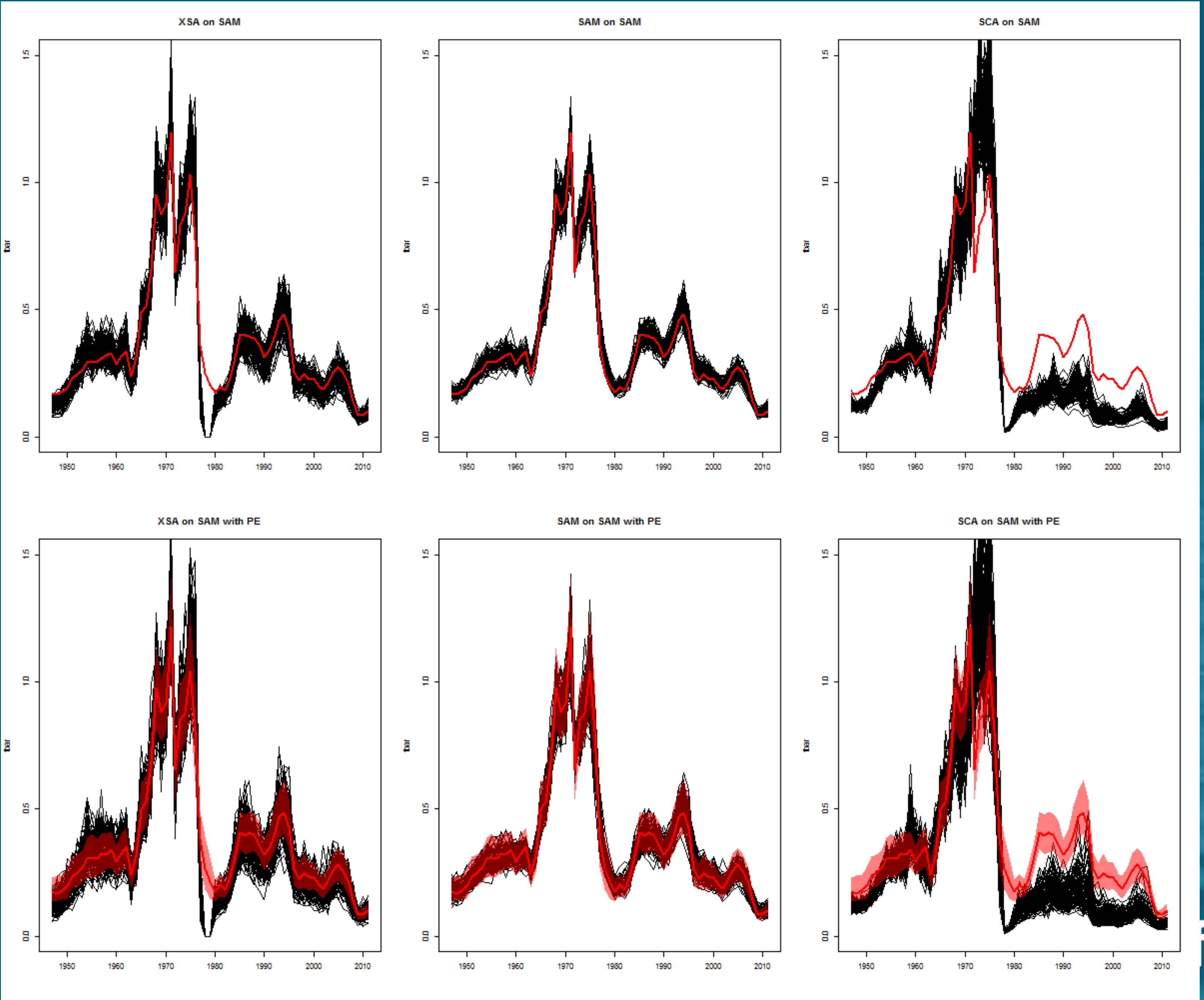
fas

North Sea Herring - Fbar

XSA on... SAM on... SCA on...

...data generated from SAM fit with obs. error

...data generated from SAM fit with obs. And process error

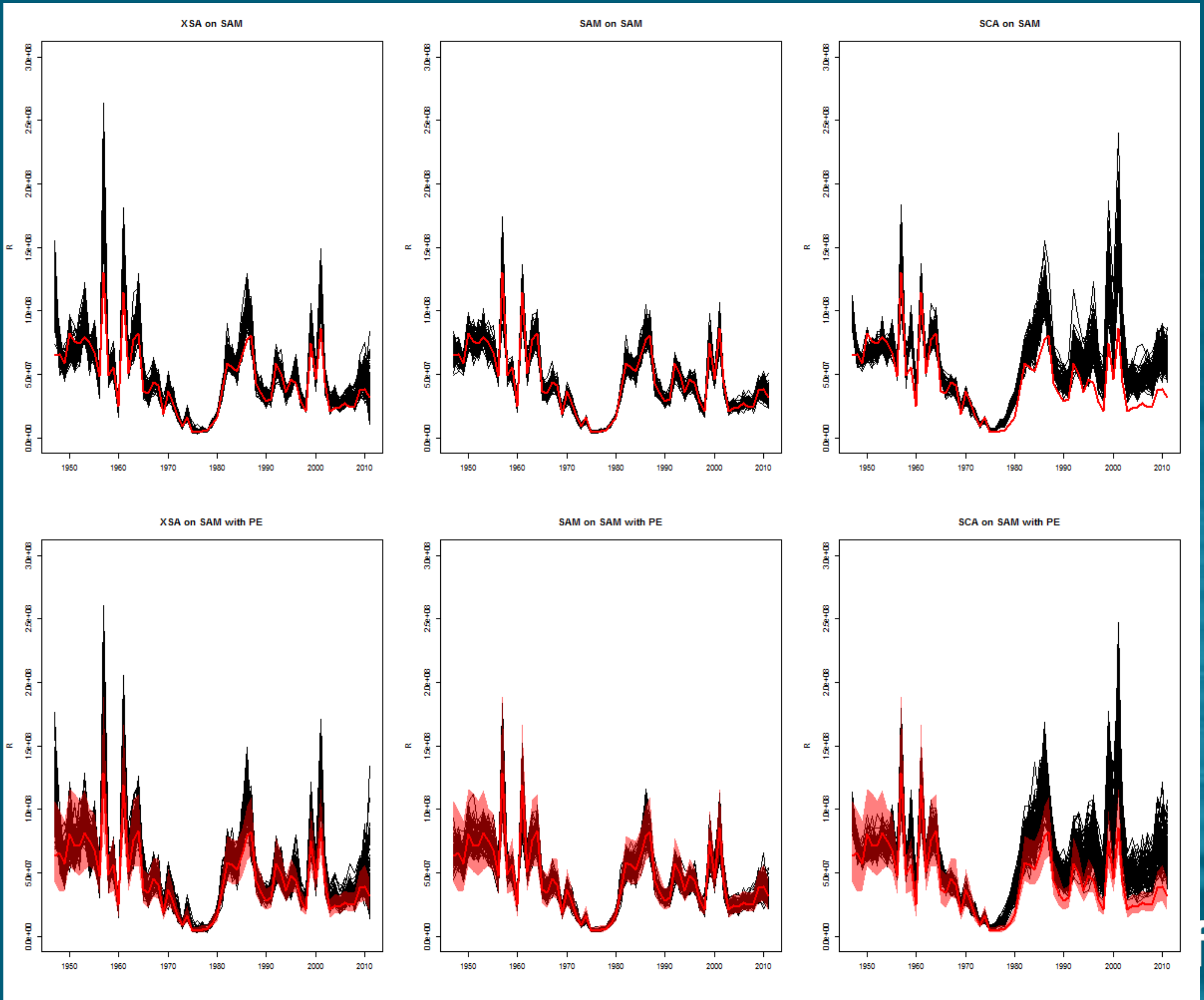


North Sea Herring - Recruitment

XSA on... SAM on... SCA on...

...data generated from SAM fit with obs. error

...data generated from SAM fit with obs. And process error

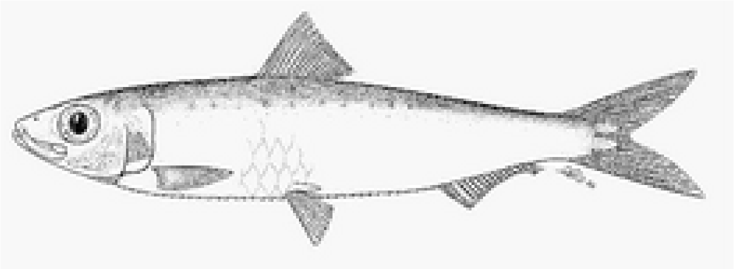


fas

Summary

- Can be a check of self-consistency of the model
- Only evaluating model similarity
- Can provide another perspective on a model fit and behaviour
- Does process error add any more information about the models than simply obs. error?

Cefas

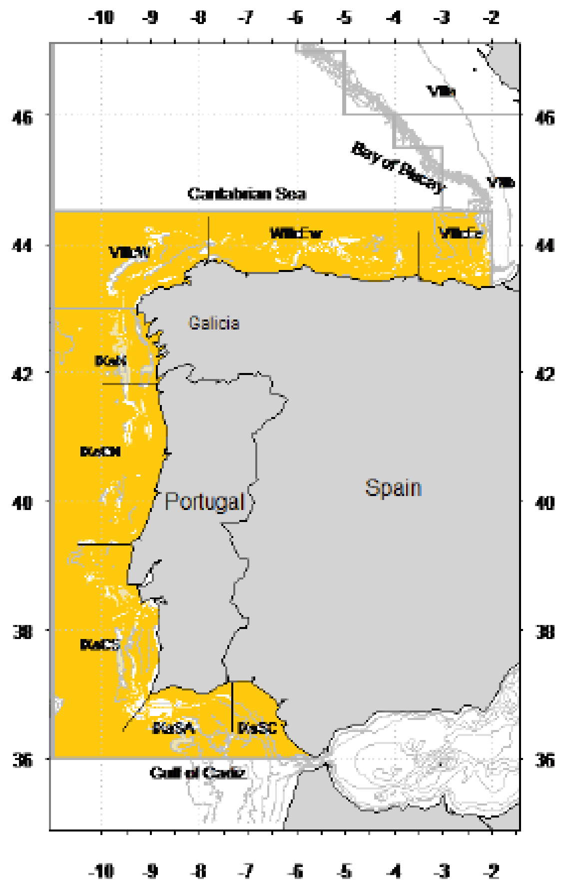


Iberian Sardine

Paul Spencer[1], Cristian Canales[2], and Jim Ianelli[1]

[1]Alaska Fisheries Science Center, Seattle, USA
[2]Instituto de Fomento Pesquero, Valpariso, Chile

General notes
(from WCSAM documents)



Current assessment model: SS3

Special Features

1) short to moderate longevity

2) pelagic

3) recruitment pulses with some regularity

4) time-varying fishery selectivity

5) Decline in recent years

Brief description of integrated catch at age models

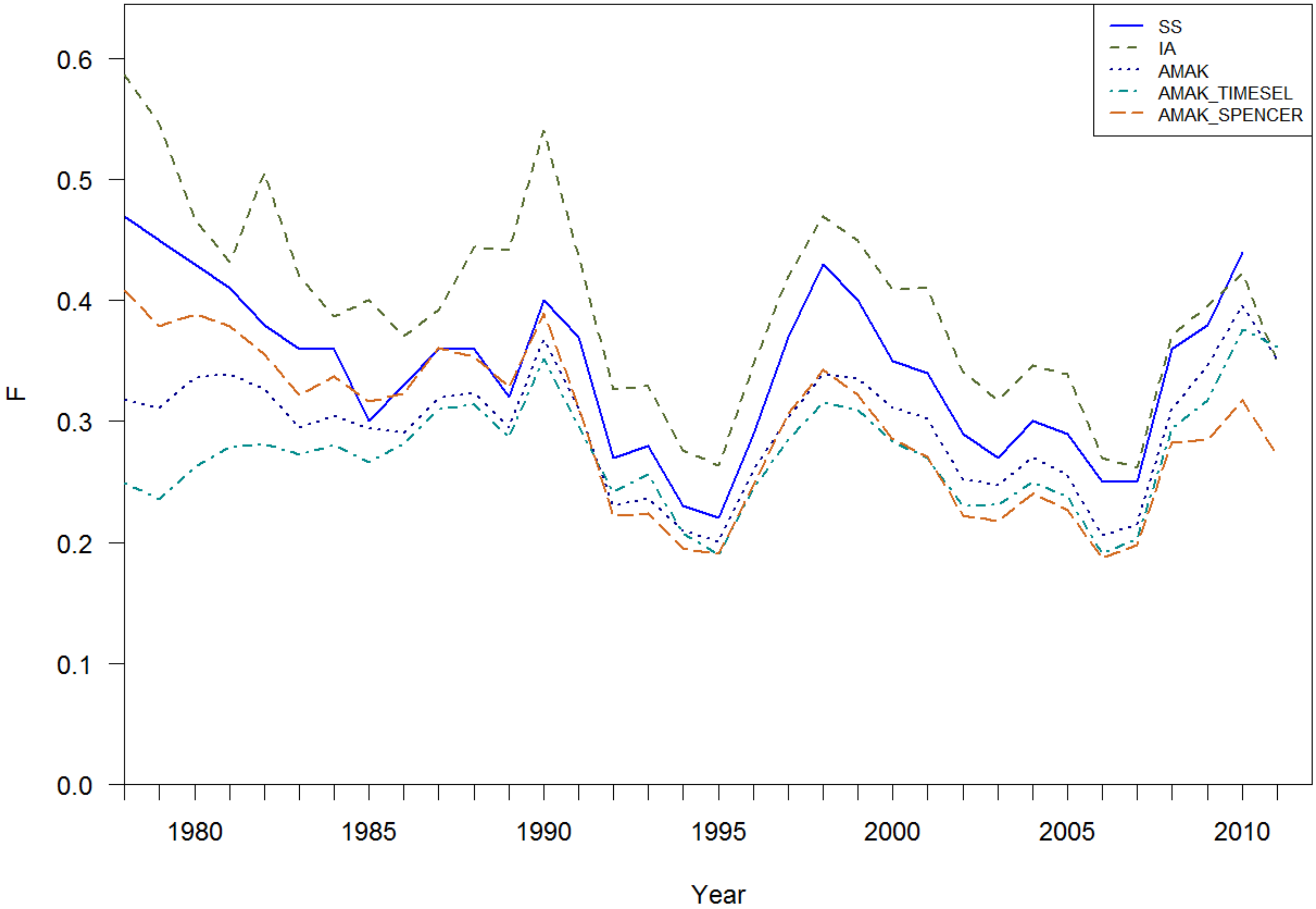
(with differences from Stock Synthesis shown)

	Model feature						
Model	**Recruitment**	**Initial N at age**	**Fish Selectvity form**	**Fish Selectivity Blocks**	**Survey Selectivity form**	**survey Selectivity blocks**	**M**
Stock Synthesis	Deviations from mean	Estimated as parameters	Random walk over ages, with ages 3-5 identical	Annually varying from 1978-1990, constant afterwards	Random walk over ages, with ages 2-5 identical	constant over time	age 1=0.8 age 2= 0.5 age 3=0.4 ages>+4 = 0.3
Spencer AMAK				**Constant over tme**			
Canales IA			**Dome shaped (ages 3-5 identical)**		**Logistic**		
Ianelli AMAK_timevar	**BH, with steepness =0.8**		**Dome-shaped**	**Blocks of constant length**	**Dome-shaped**		
Ianelli AMAK	**BH, with steepness =0.8**		**Dome-shaped**	**Constant over tme**	**Dome-shaped**		

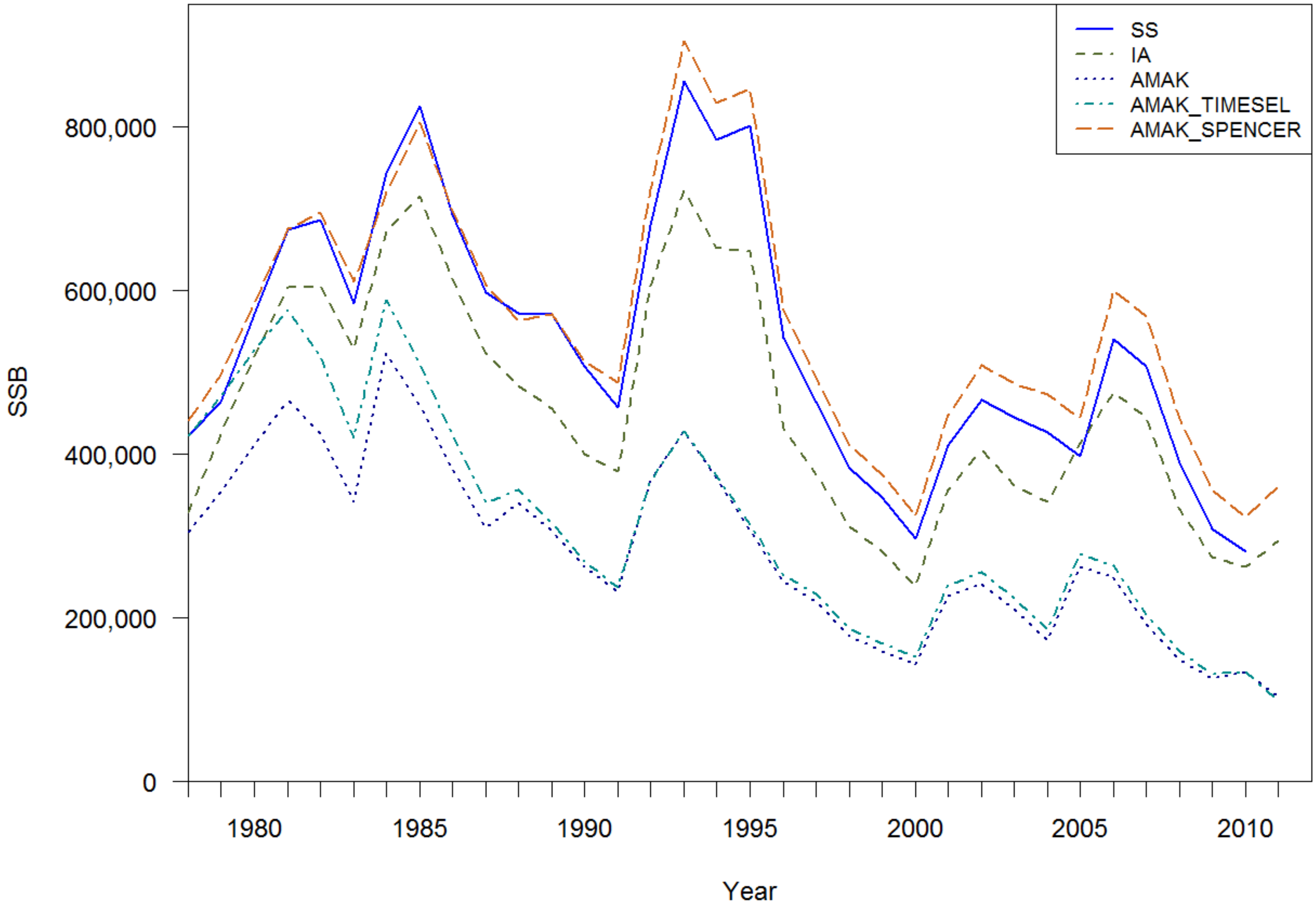
Data

	Fishery	Acoustic Survey (Abundance)	DEPM survey (SSB)
Landings	1978-2011		
Survey Index		1996-2003 2005-2011	1997, 1999, 2002, 2005, 2008, 2011
Catch at age	1978-2011	1996-2003 2005-2011	
Weight at age	1978-2011	1996-2003 2005-2011	
Maturity Ogive			2002, 2005, 2008, 2011
Age range	0-6+	1-6+	

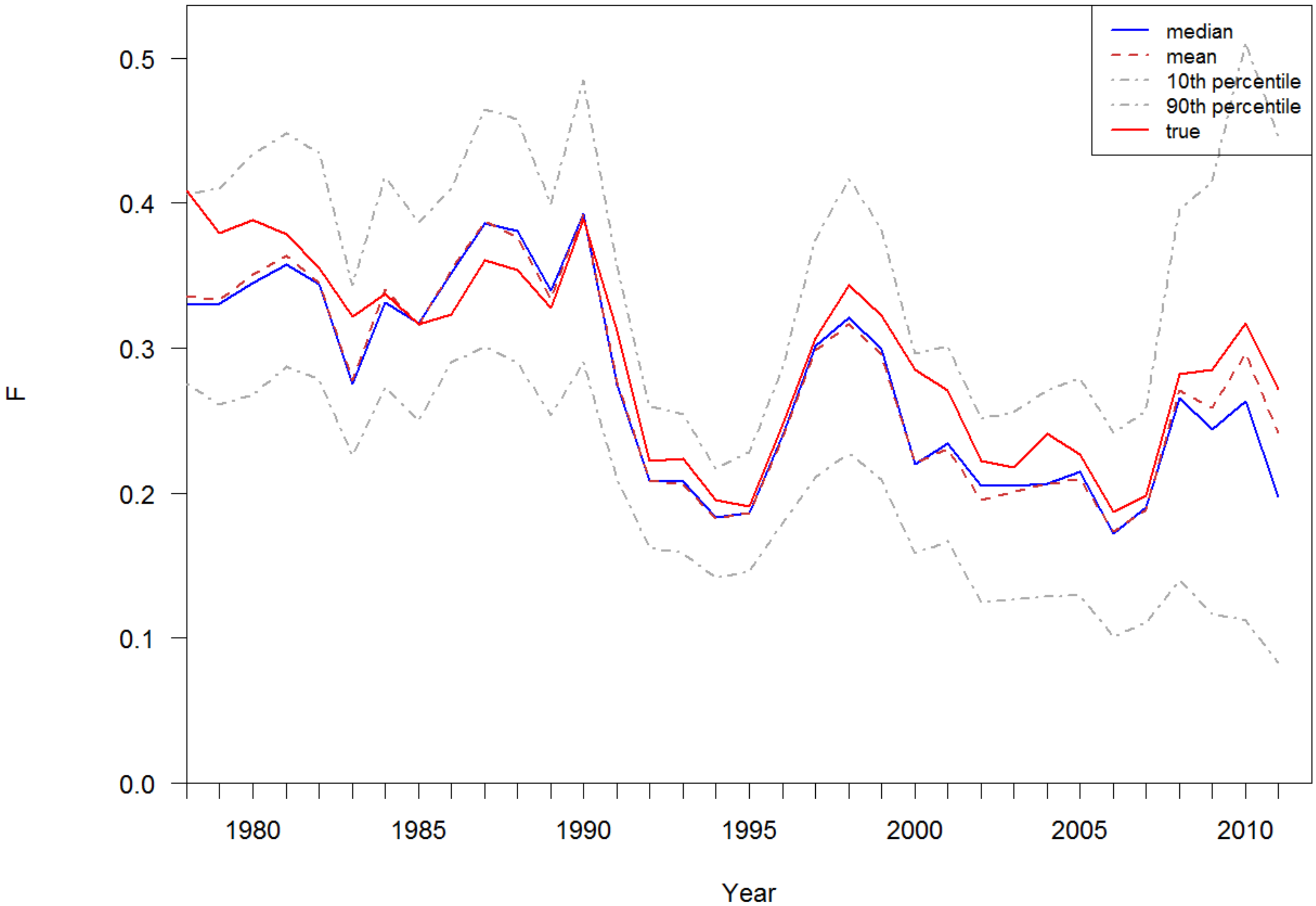
F Estimates of all models fit to real data



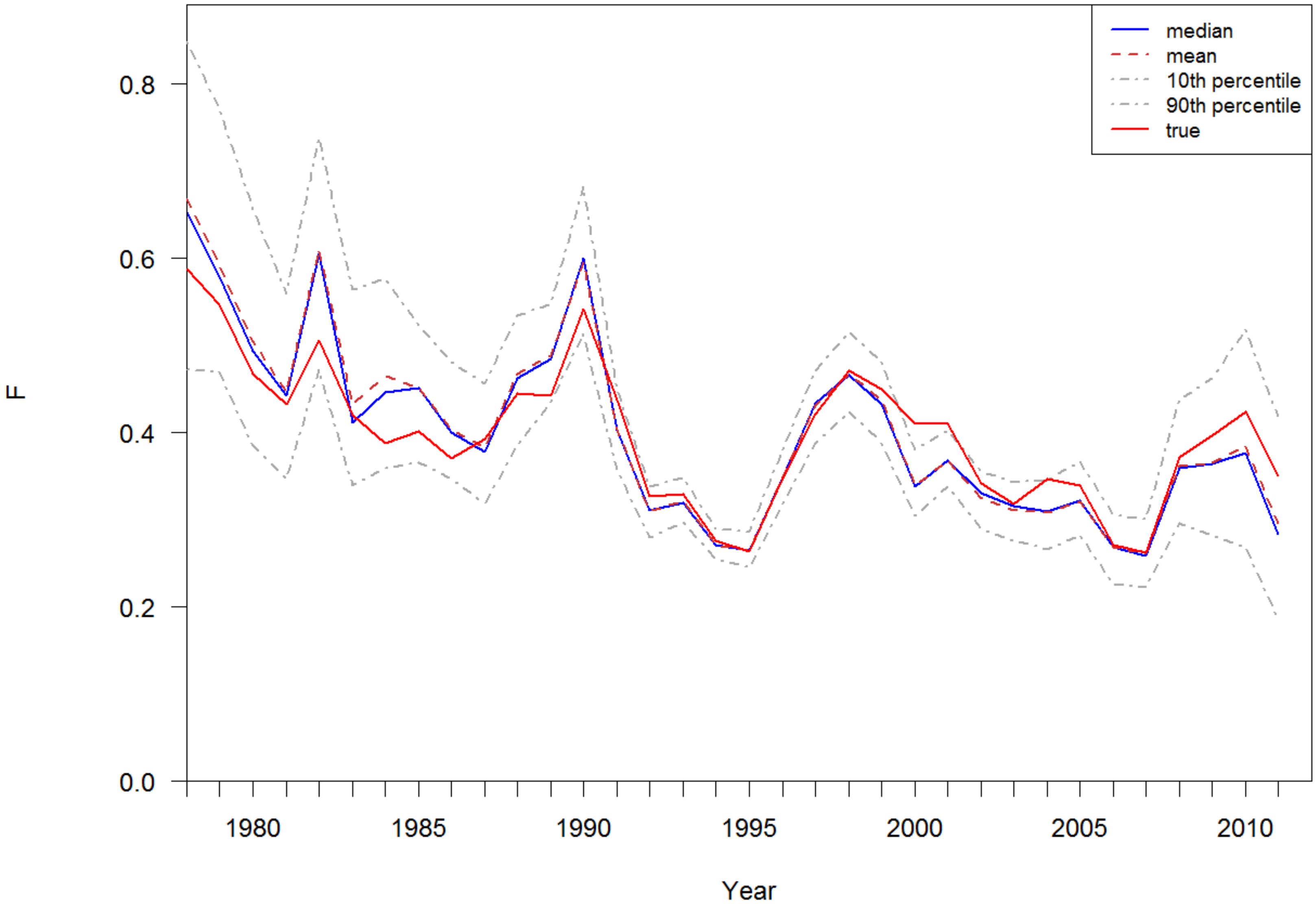
SSB Estimates of all models fit to real data



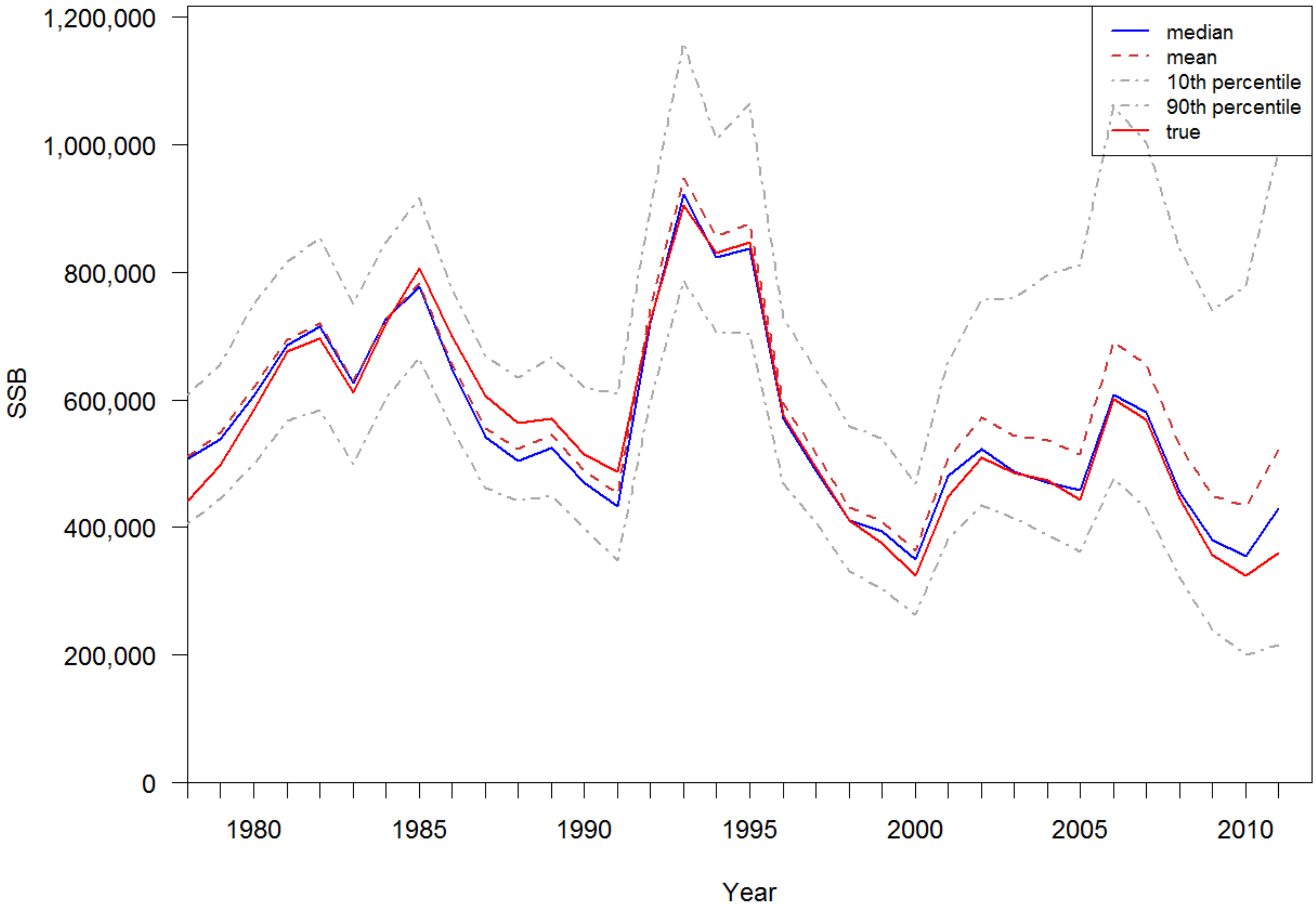
F Estimates AMAK_Spencer 'self-test'



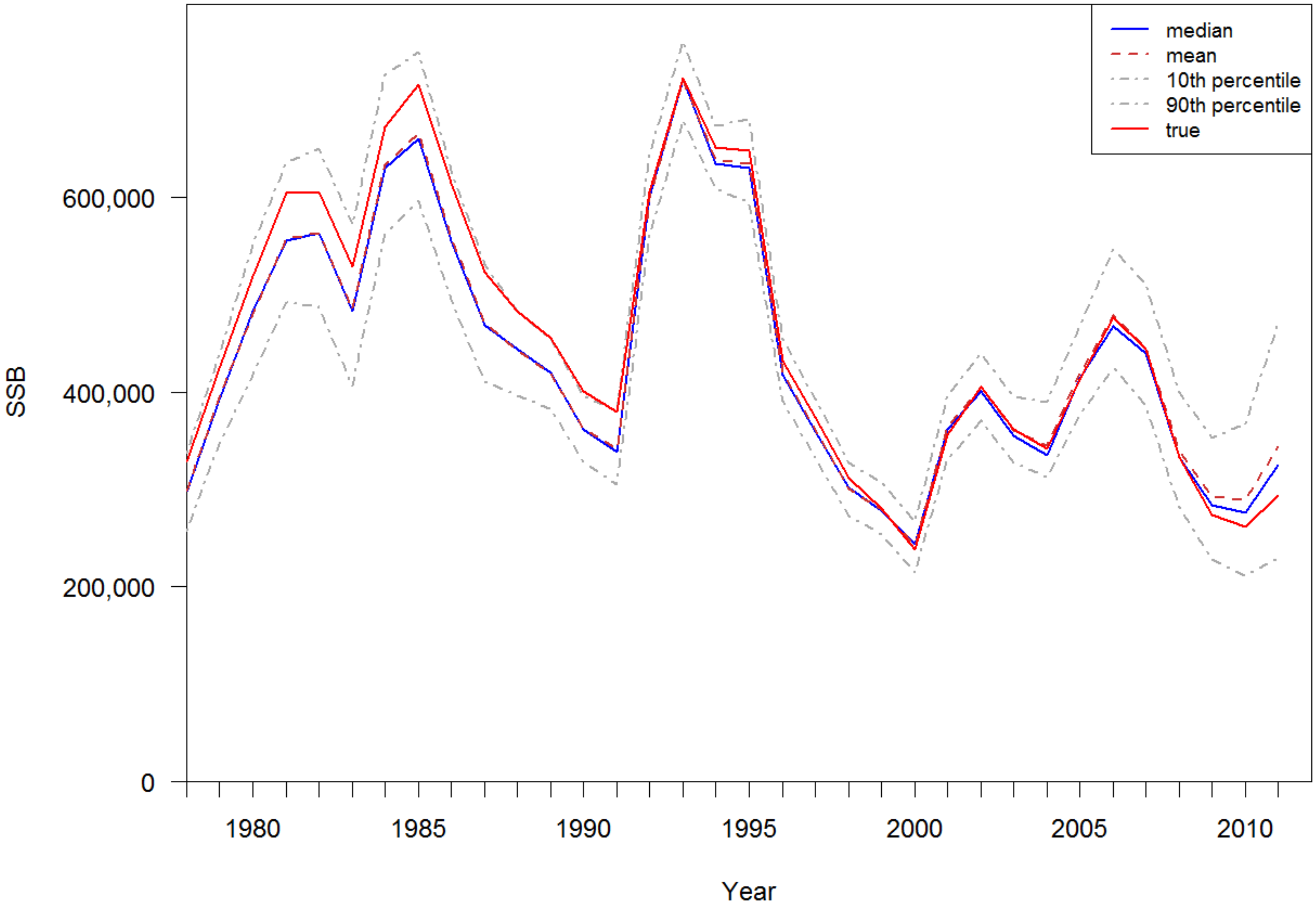
F Estimates from IA_Canales 'self-test'



SSB Estimates from AMAK_Spencer 'self-test'

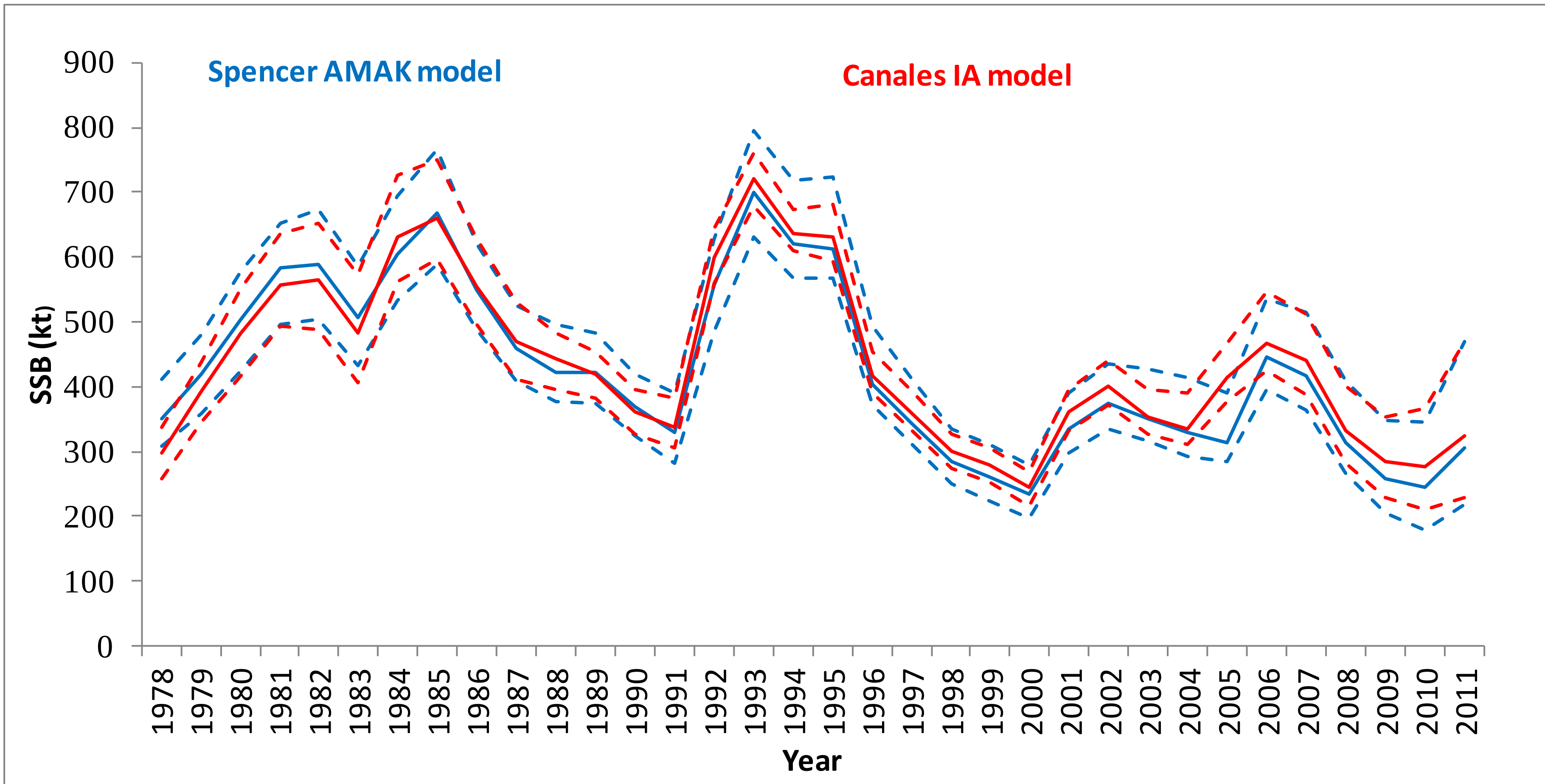


SSB Estimates from IA_Canales 'self-test'

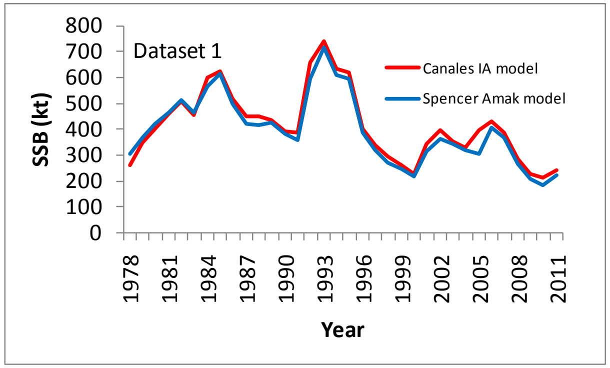


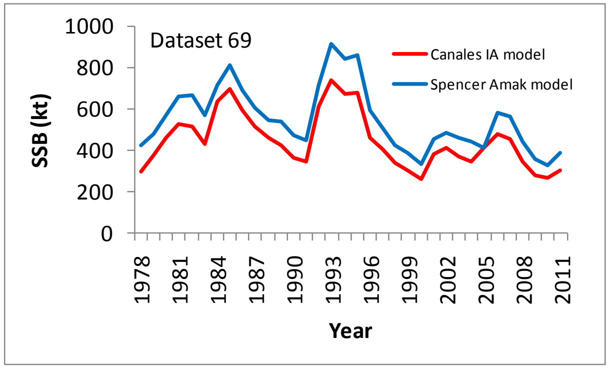
Cross-test

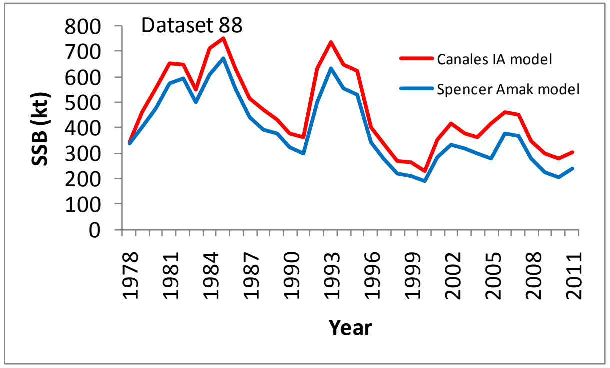
Each model fit to the data generated from Canales IA model
Medians, and 10th and 90th percentiles



Results for individual runs will vary







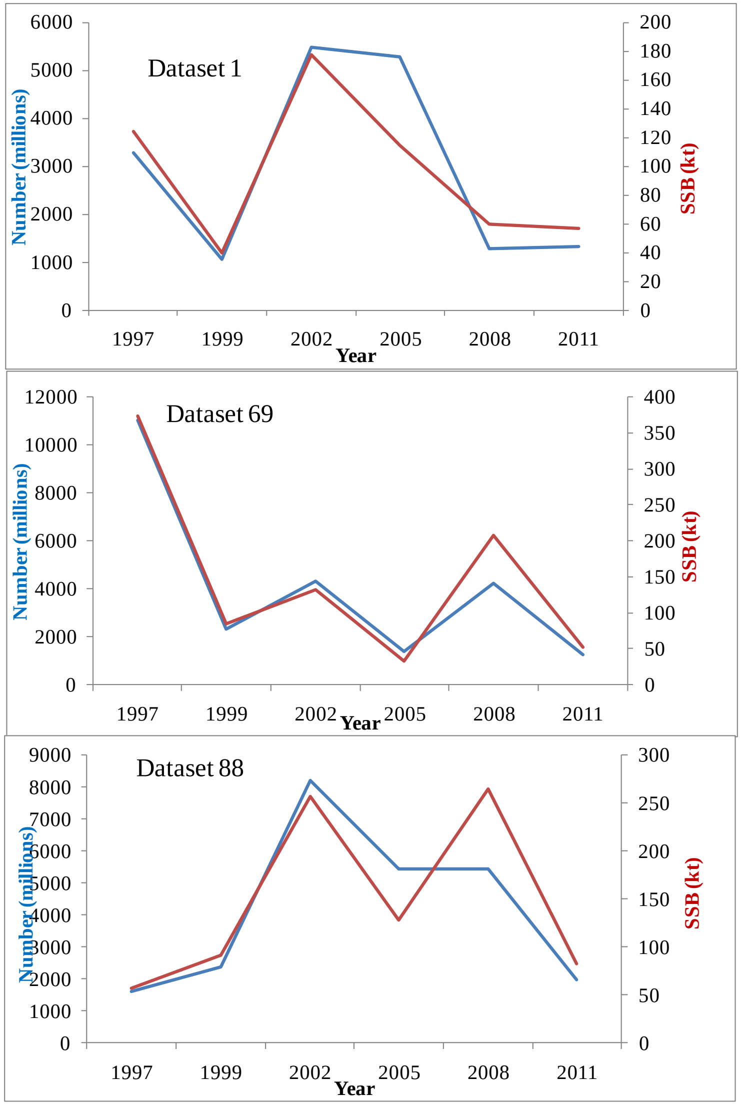
Difference in interpretation of simulated DEPM survey index

Simulated as population numbers, after some attempts to get simulated SSB.

Canales – modeled as population numbers

Spencer – Used the simulated survey numbers as if they were SSB, with difference in numerical scale to be absorbed by the survey catchability.

This may have contributed to the differences between the fits to individual data sets



Examples of difference between DEPM numbers and SSB

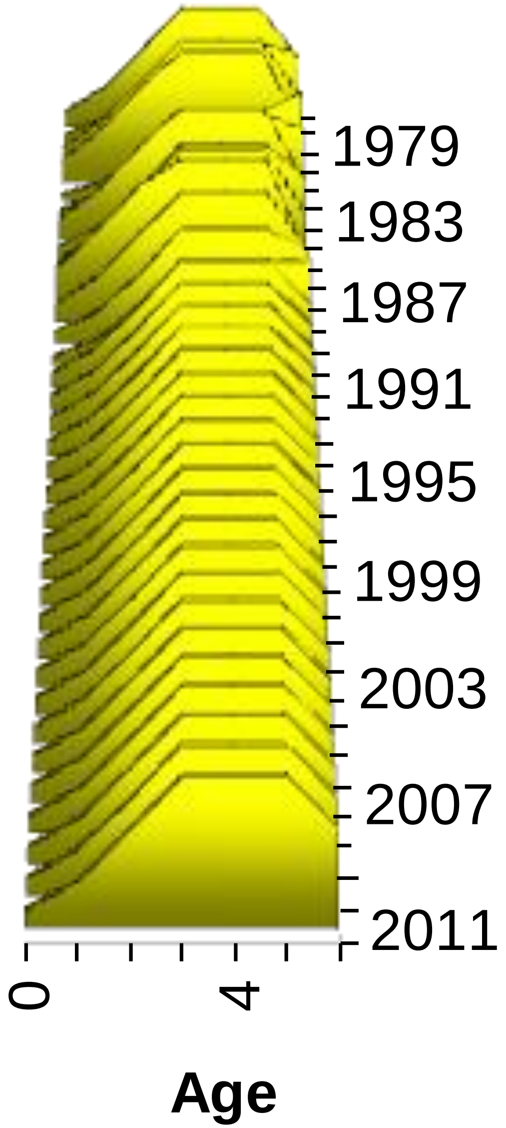
These are conversions of the simulated DEPM numbers to SSB, using the data on weight and proportion mature at age.

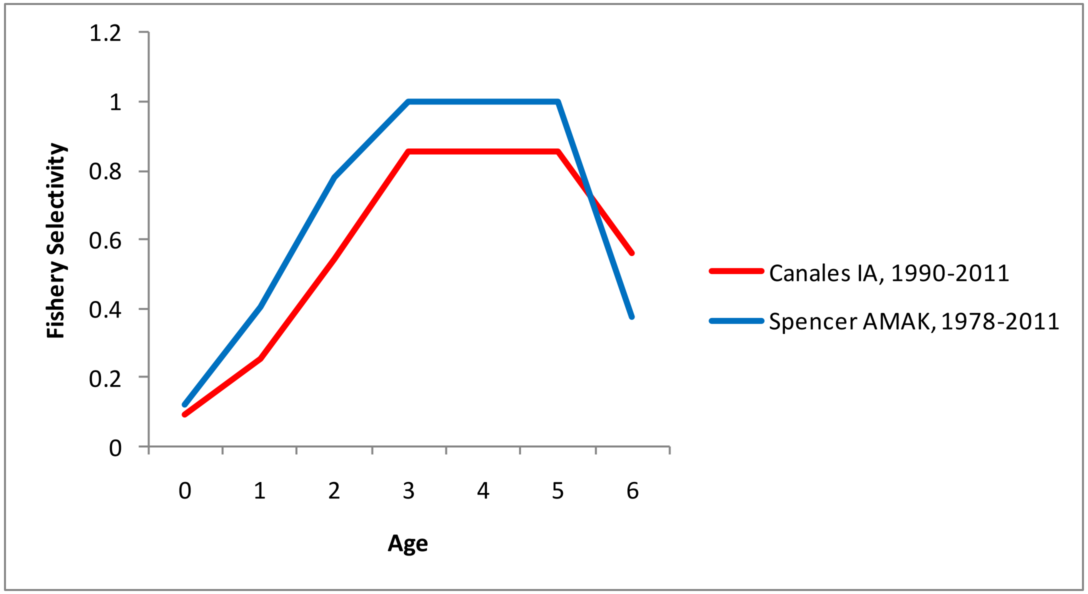
Point for discussion

Dome-shape selectivity identified in initial WCSAM documents

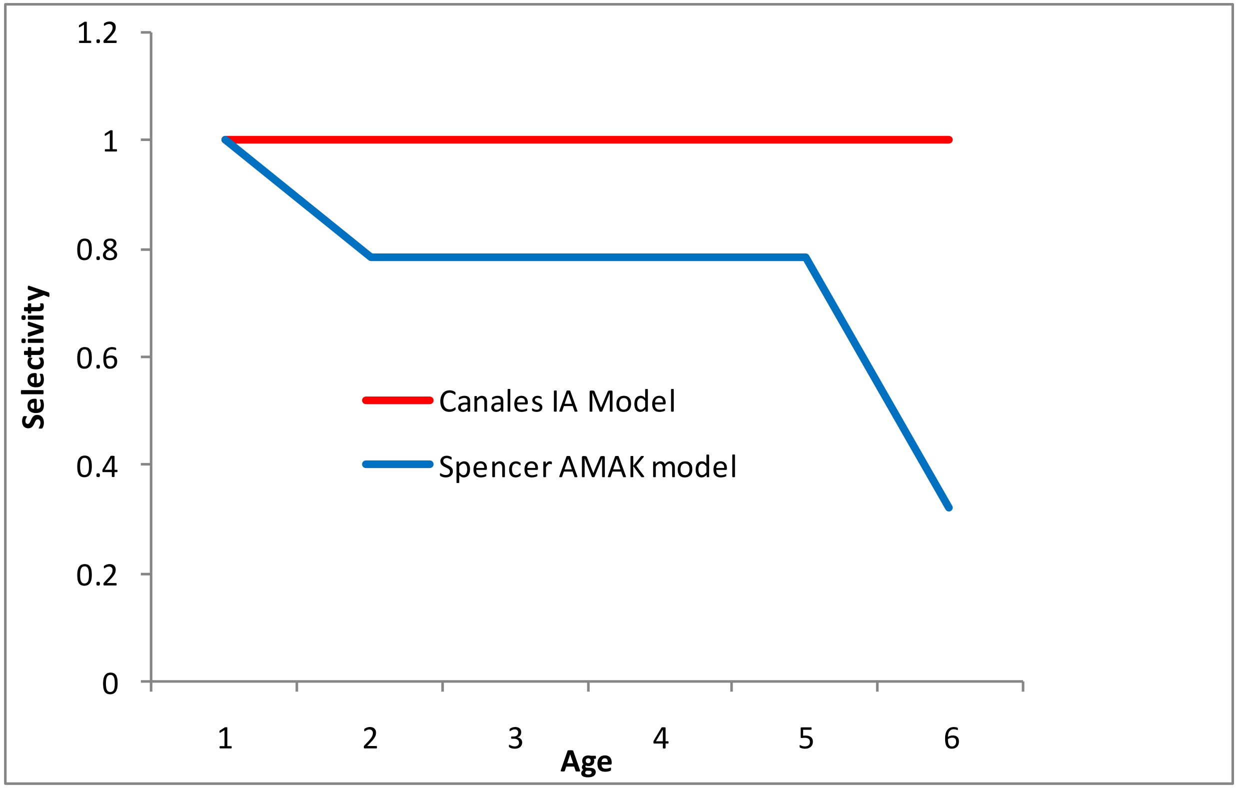
Fishery selectivity curves

Canales IA model





Survey Selectivity curves



Potential mechanisims on selectivity (from WKEPLA_2012)

Ages 2 and up are similar in size, so differences in selectivity between ages are not likely due to size

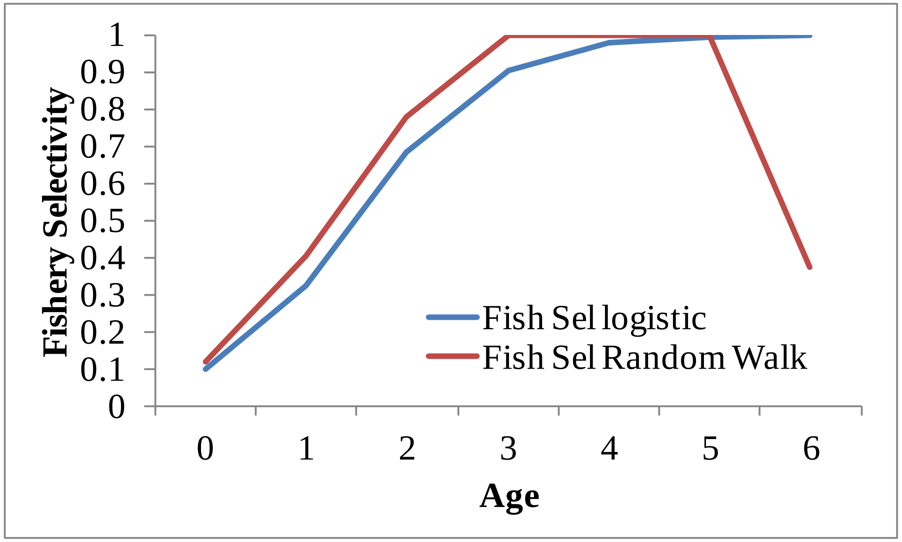
Small segregate closer to shore, but no spatial segregation of fish larger than ~ 23 cm. (ages 2+)

Low fishery selectivity of ages 0 and 1 could be lack of market interest.

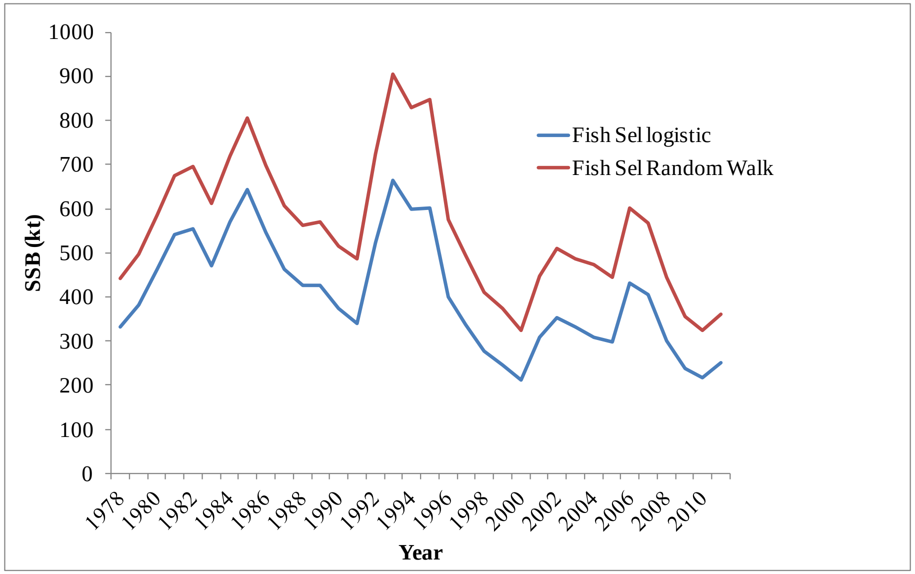
Declines in selectivity for ages 6+ in both fishery and survey

". . .the causes are still uncertain; differences in availability or senescence seem unlikely but the hypotheses that older fish are less catchable cannot be excluded"

Fit of AMAK model assuming logistic fishery selectivity



	- ln likelihood	DEPM survey q	Acoustic survey q
Random_walk	164.87	0.86	1.61
Logistic	171.12	1.26	1.94



Initial conclusions; further Points for Discussion

1) The models applied to this stock were broadly similar Integrated Assessments, and generally produced similar results (on average).

2) The different *survey* selectivity forms, and of time-varying *fishery* selectivity, between the Spencer and Canales models seems to have little effect on model output (on average).

3) Differences between the Spencer and Canales models in fits to individual datasets may reflect interpretation of the DEPM index.

4) The two Ianelli AMAK models also indicate relative little influence from time-varying fishery selectivity.

5) Results are sensitive to form of the fishery selectivity curve, and emphasize the important of understanding the mechanisms.

North Sea Cod

Fits to Real data

WCSAM workshop, Boston July 2013

José De Oliveira

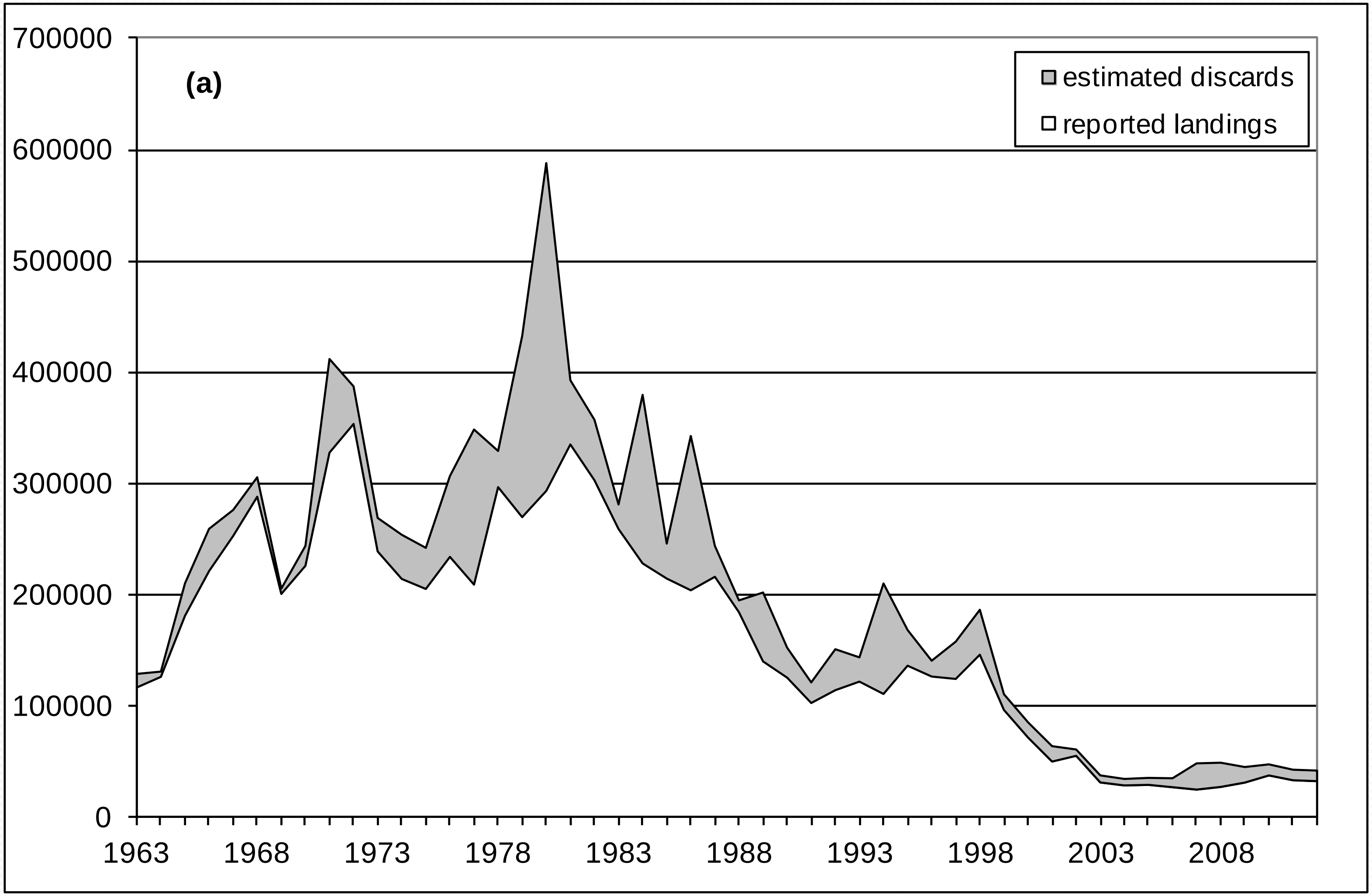
Data available for simulation exercise

Data:
Landings and Discards: 1963 – 2011, ages 1 – 7+ (use total catch only)
IBTS Q1 survey: 1983 – 2012, ages 1 – 5

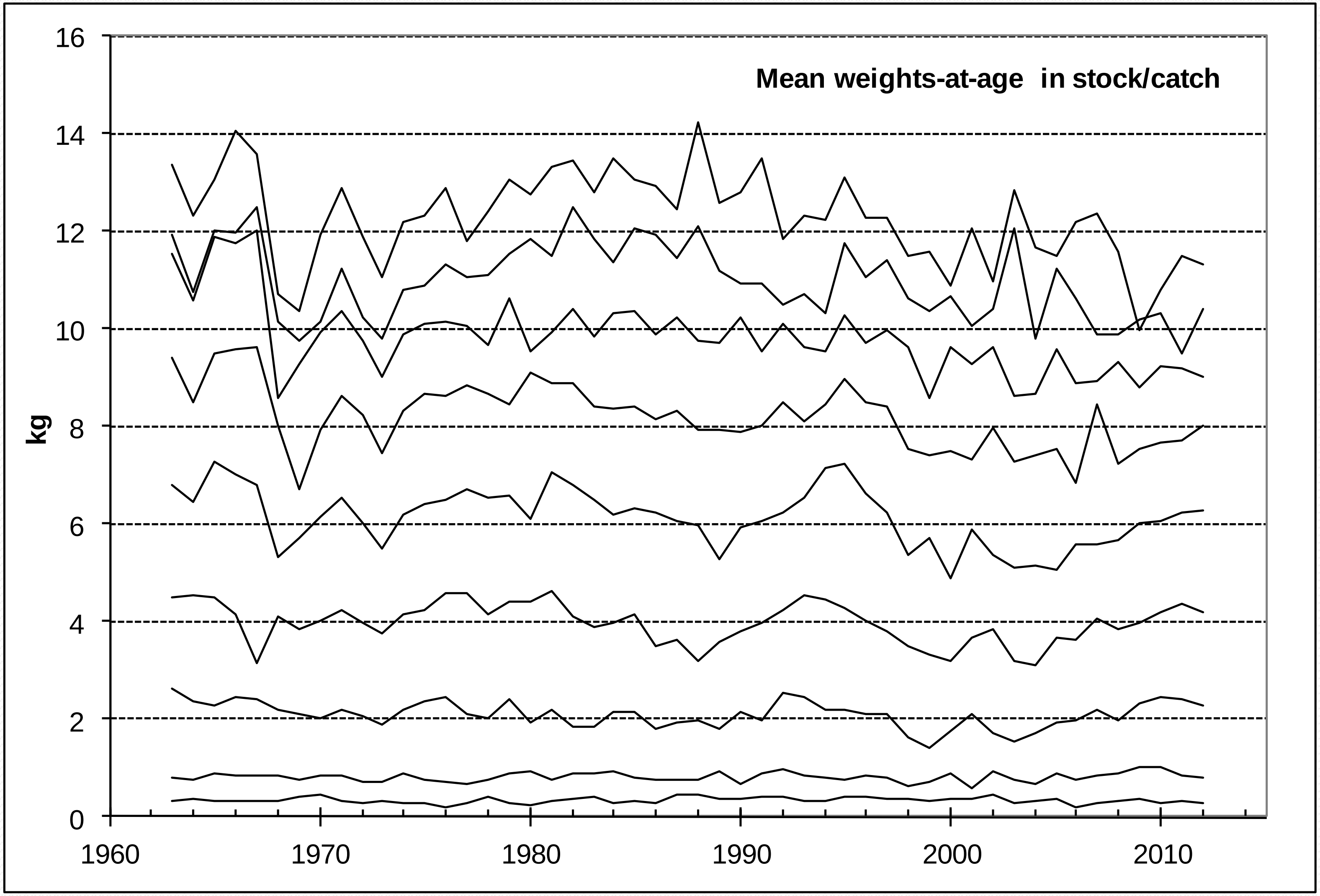
Natural mortality from multispecies model, varies by age and year
Maturity from surveys, assumed constant over time
Mean weights-at-age in the catch (landings & discards; =stock weights)

Assessment:
SAM (State-space Assessment Model) with catch scaling (1993-2005)

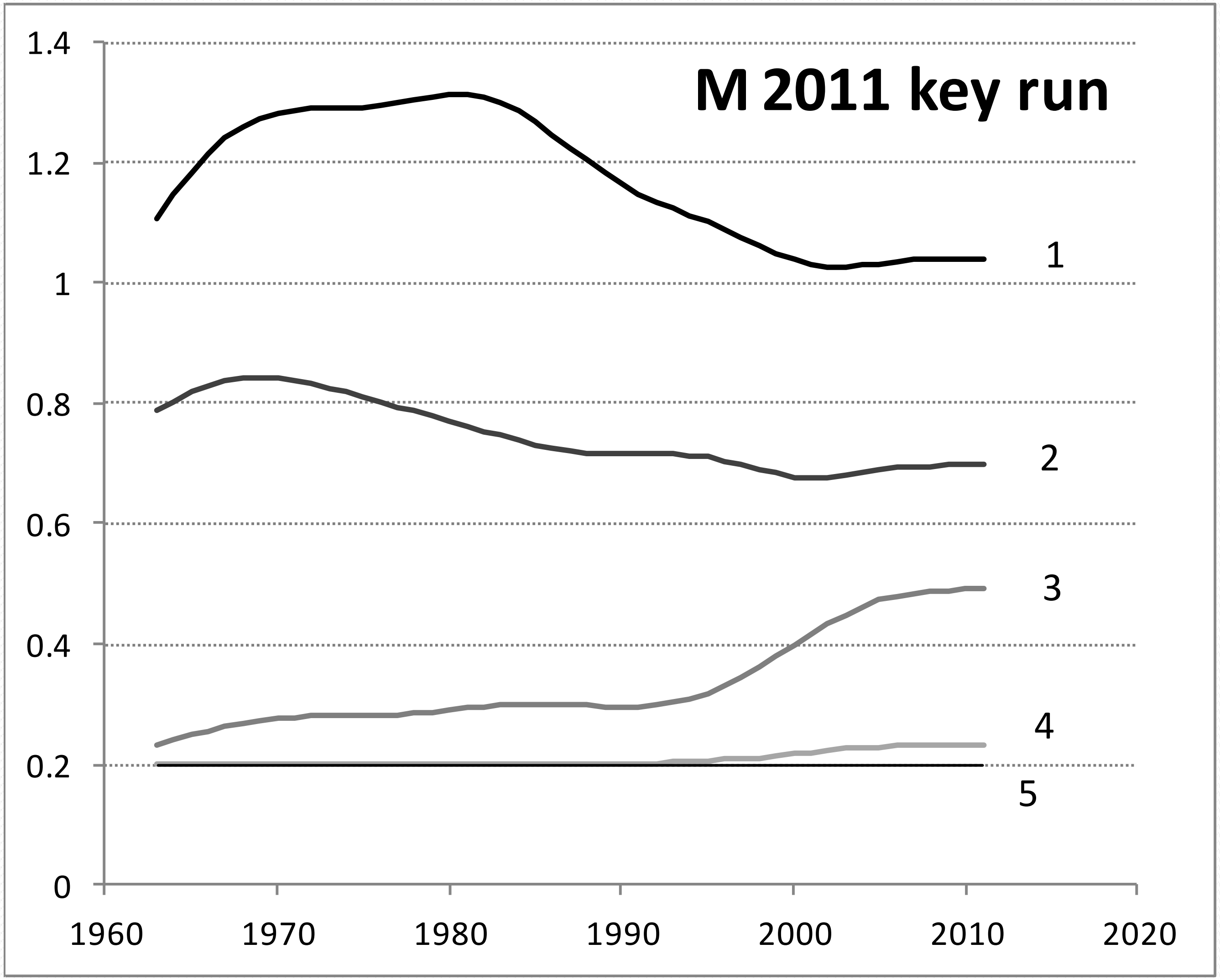
Landings and Discards



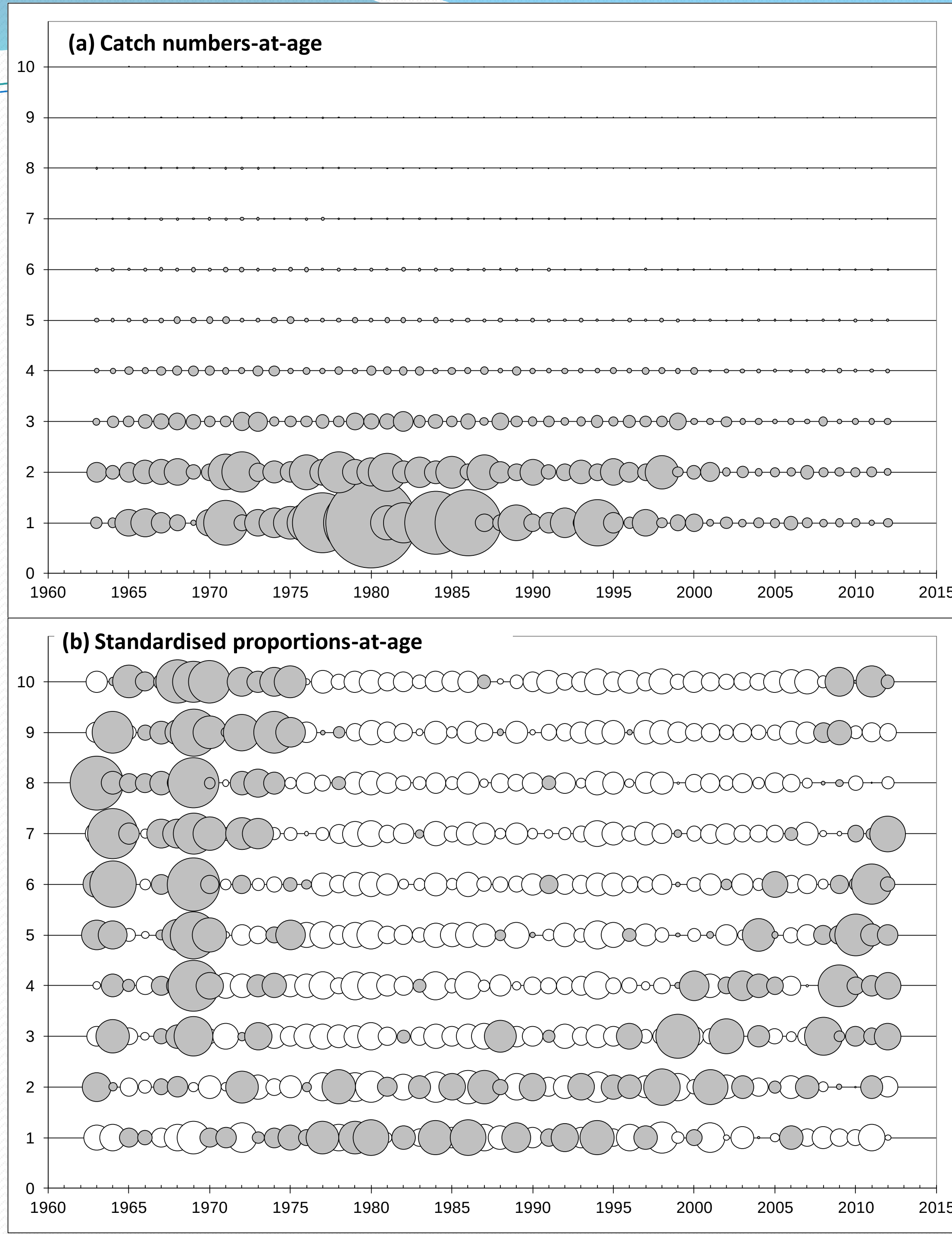
Mean weights in Catch/Stock



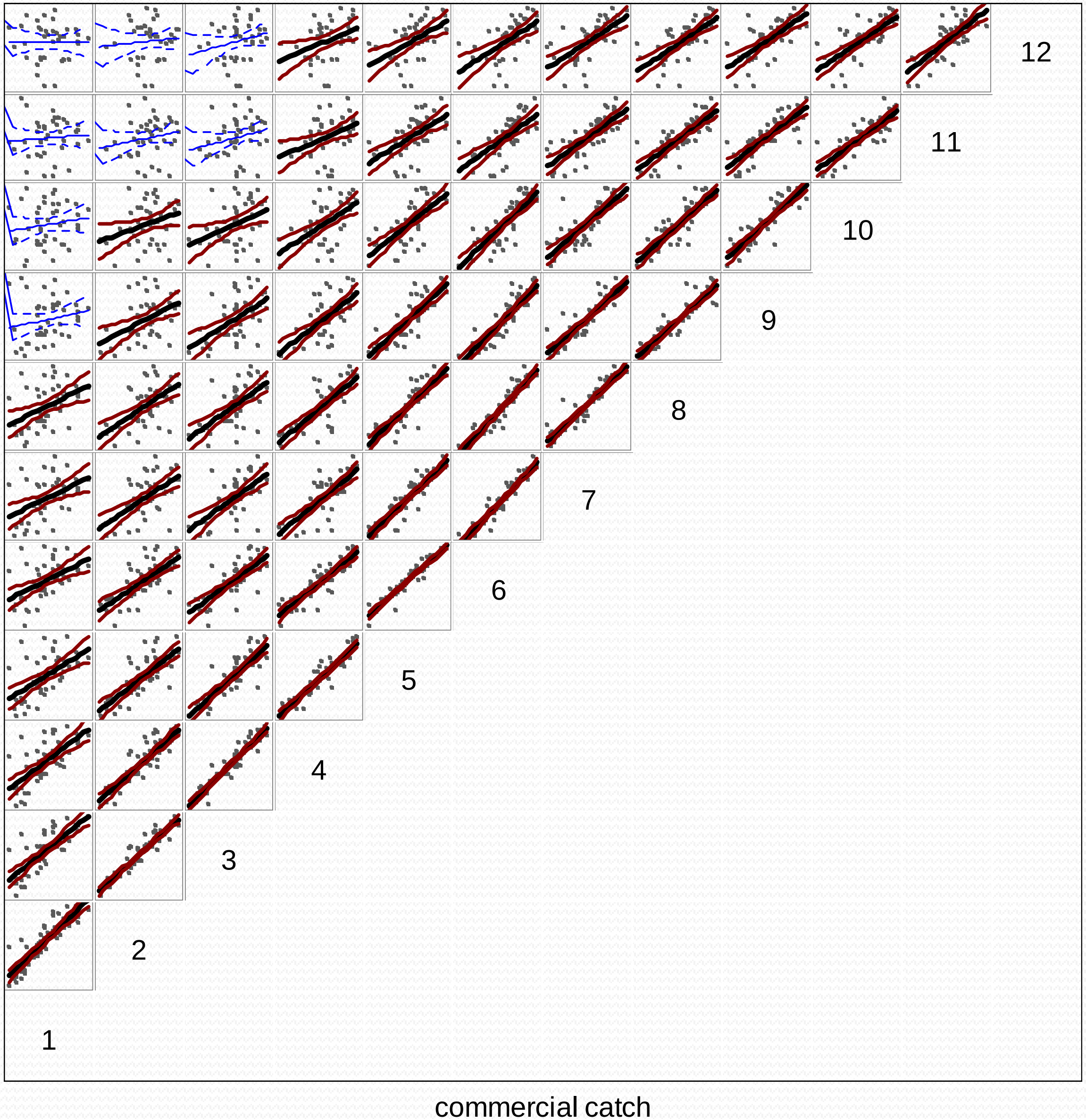
Natural Mortality



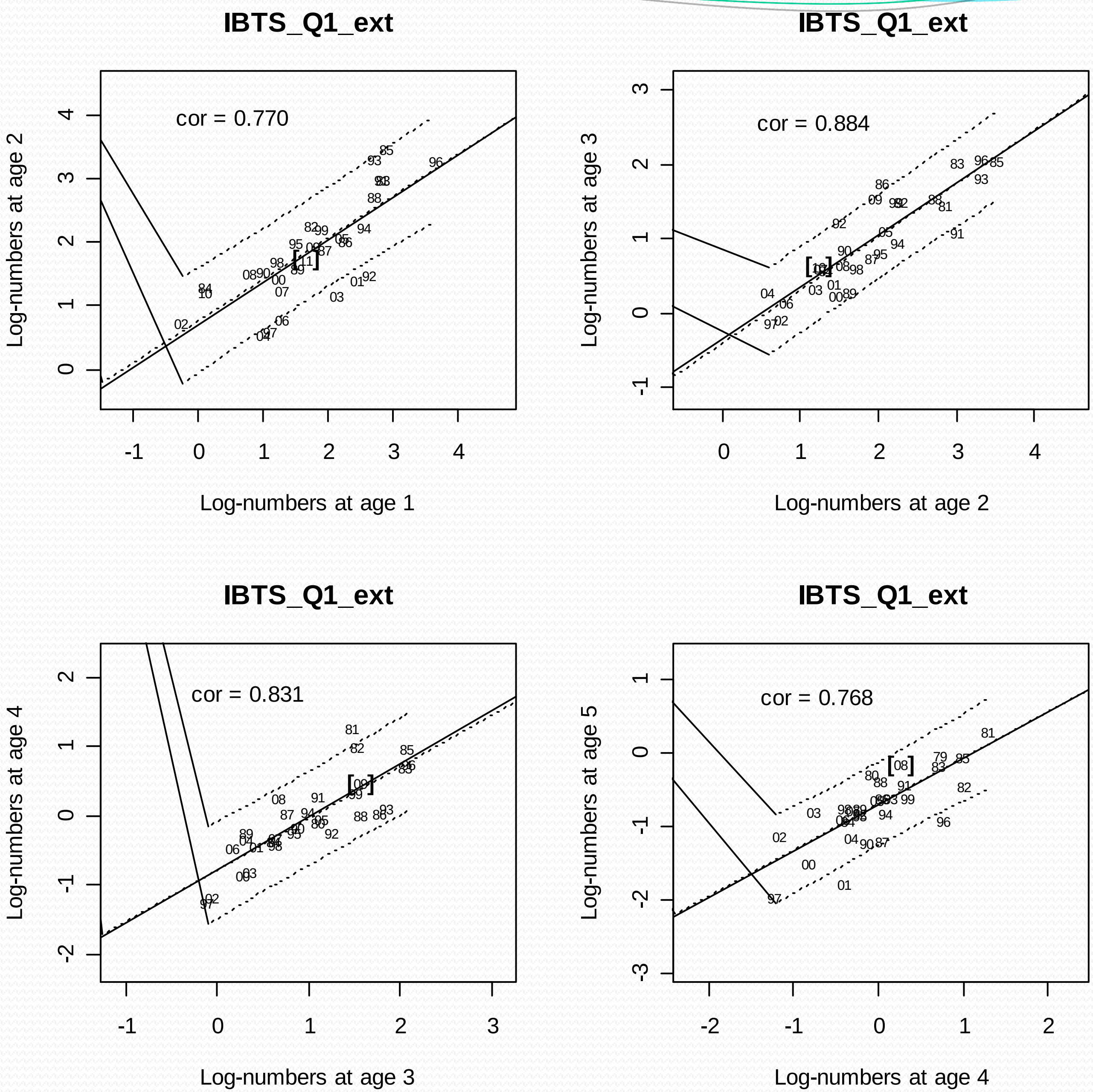
Catch-at-age matrix



Catch-at-age correlations

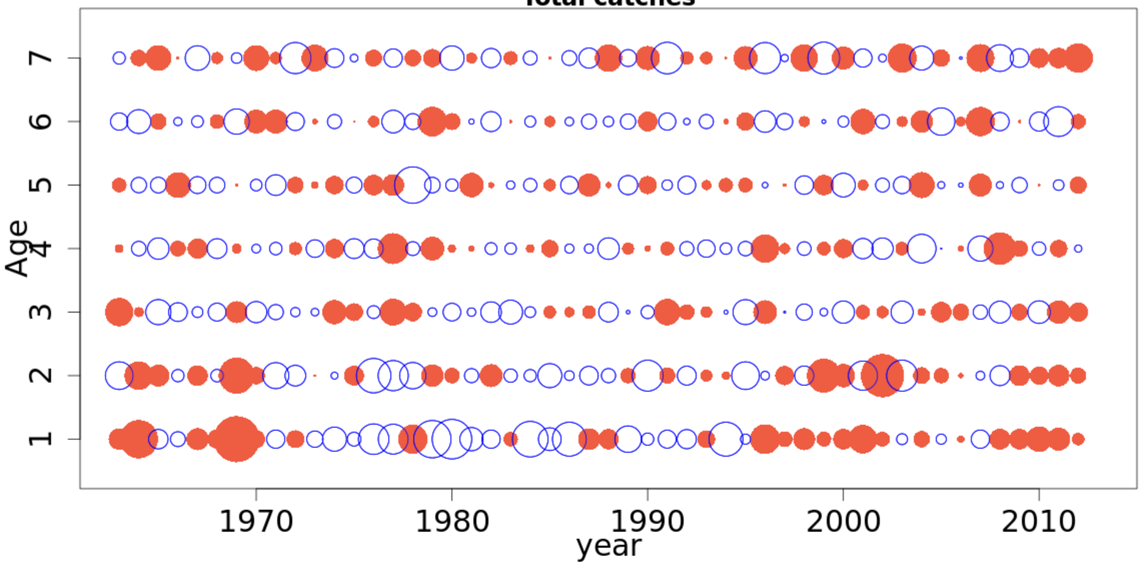


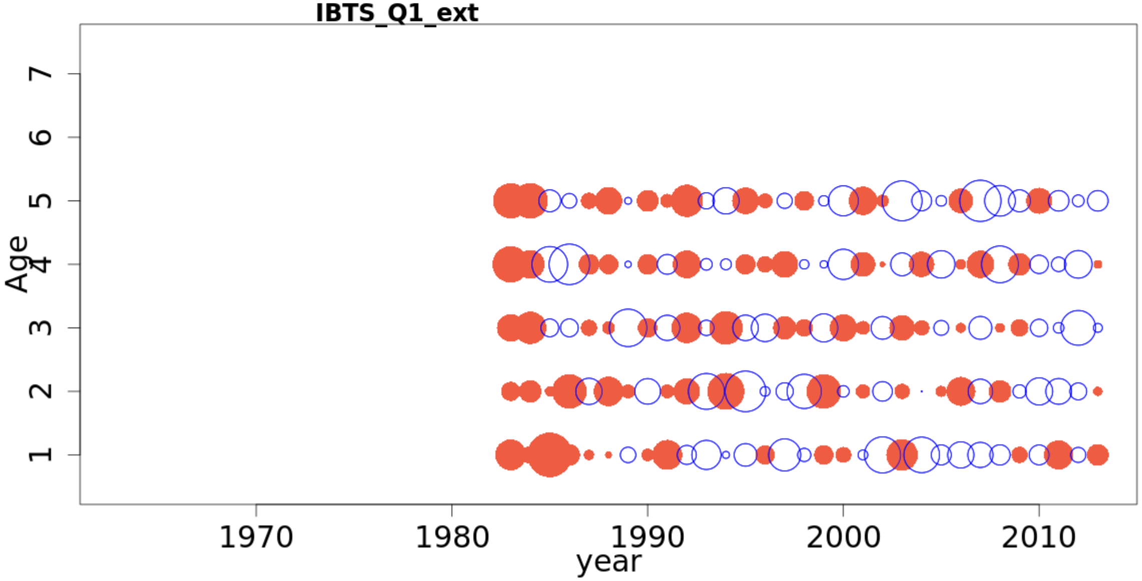
Within-survey consistency: IBTSQ1



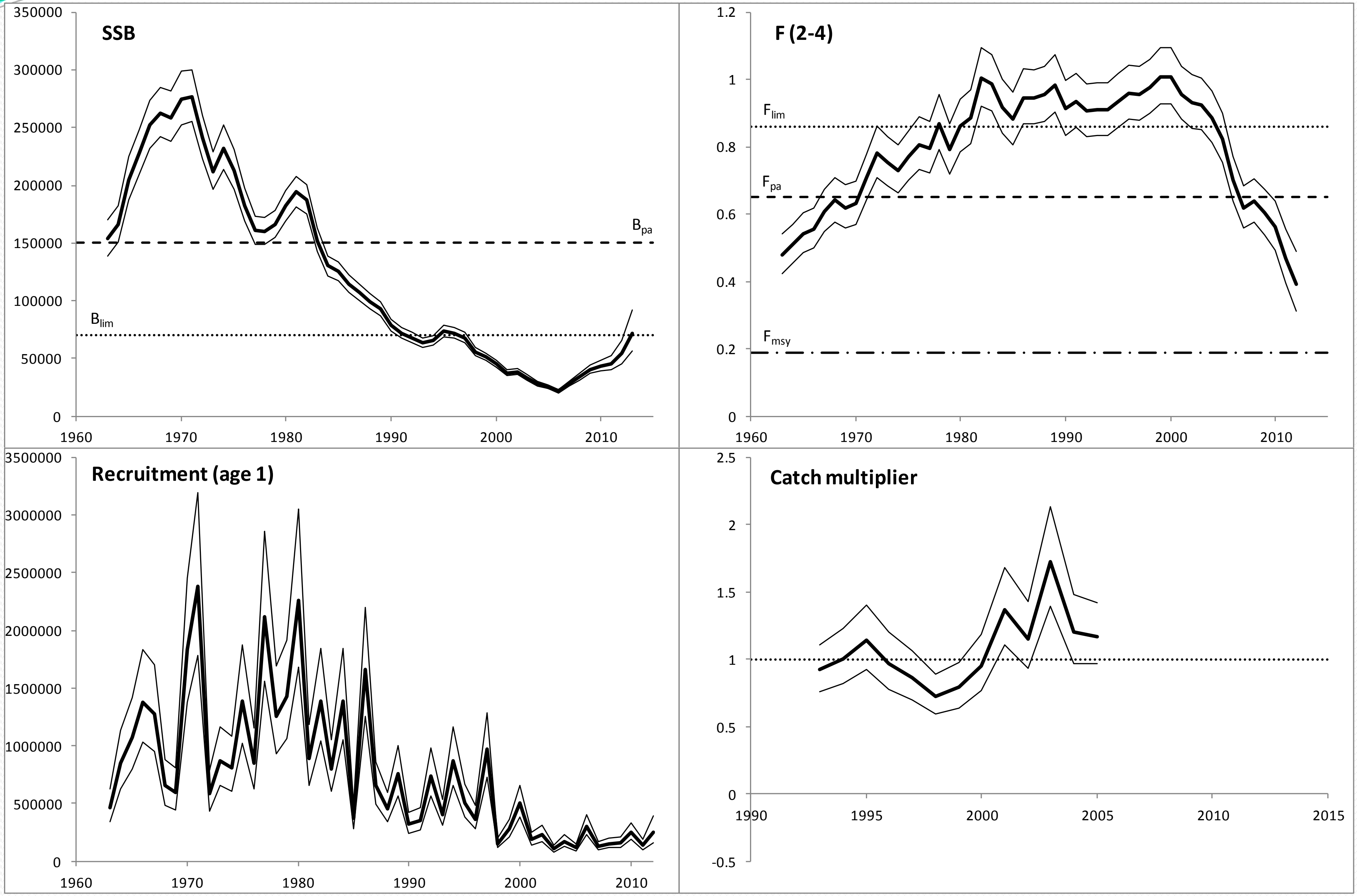
SAM Assessment

Residuals





SAM Assessment



Models fitted

SISAM Categorisation Scheme:

1. Catch-only	-
2. Time-series	-
3. Biomass Dynamic	-
4. Delay difference	-
5. Age-structured production	-
6. VPA	XSA
7. Statistical catch-at-age	SAM, Stoch-ASPM, BAM, AMAK, ASAP
8A. Integrated analysis: length	-
8B. Integrated analysis: age	SS

Models fitted

XSA

(Extended survivors analysis)

VPA-derived

F>M
catch assumed exact (no gaps)
requires tuning index

Iterative algorithm terminating when Fs converge between iterations

SAM

(State-space assessment model)

Random effects:
lnF follows random walk, correlated across ages
lnN noise term independently normal

Proc. variance: σ_F, σ_R, σ_S
Obs variance: $\sigma_{C,a}$, $\sigma_{S,a}$
q_a for each index
S/R params: α, β
Corr. Param: ρ

Includes catch scaling

Stoch-ASPM

(Stochastic age-structure production model)

Estimates Fy, Ry-deviates

Selectivity-at-age in block periods, index q constant

S/R included

Penalised likelihood:
logN for total catch and index
Adjusted logN for age-comps

Includes catch scaling

Models fitted

BAM

(Beaufort assessment model)

Estimates Fy, Ry-deviates

Selectivity-at-age (param)
index q can vary

S/R included (Methot-Taylor method for constr.)

Penalised likelihood:
logN for total catch and index
Robust multinomial for age-comps
Francis (2011) method for adjusting likelihood component weights so that sdnr near 1.

AMAK

(Assessment method for Alaska)

Prior distributions on M and q

Time-varying curvature penalty allows selectivity params to change of time and age

ASAP

(Age-structured assessment program)

Can treat discards explicitly

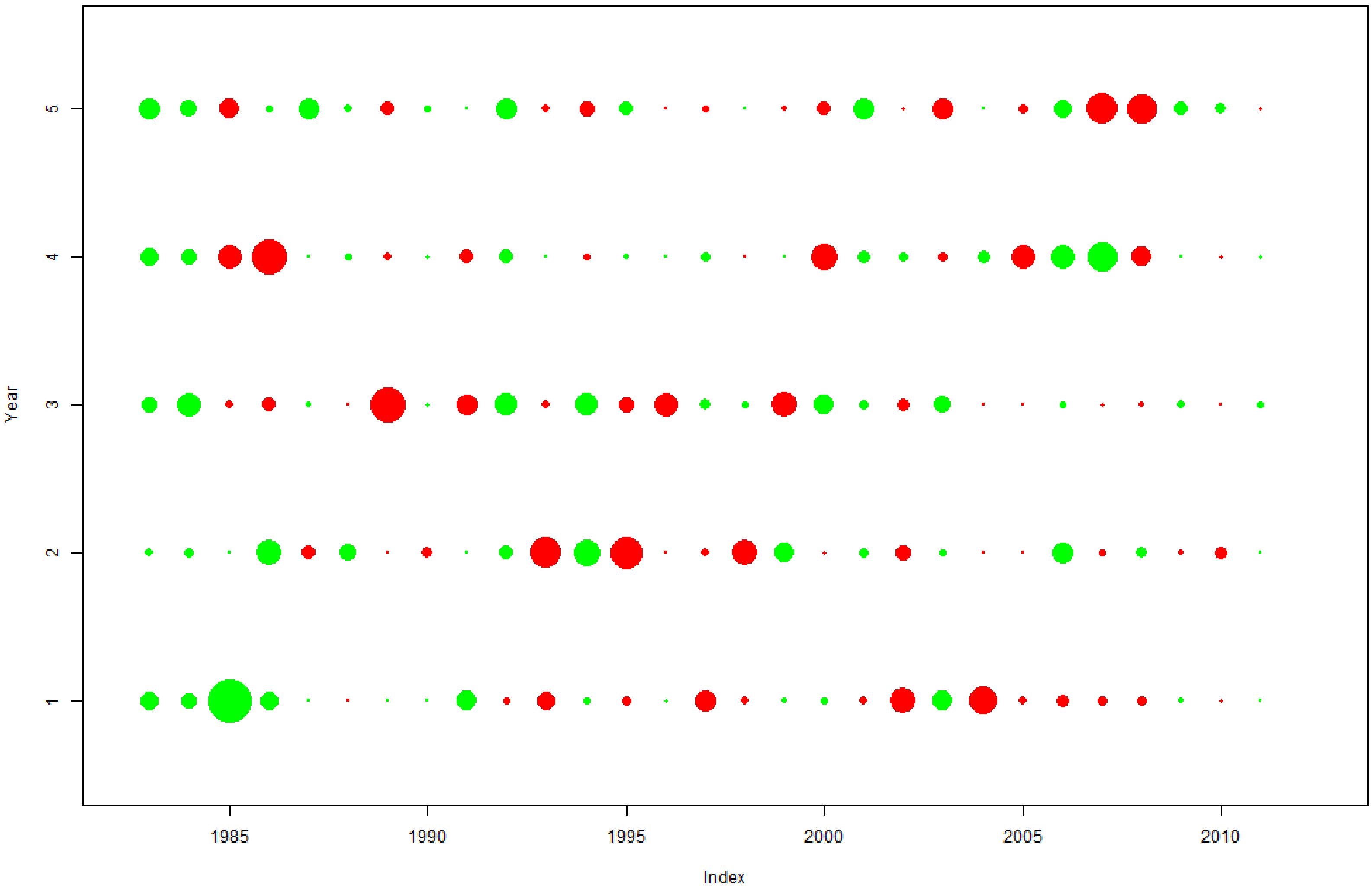
Selectivity can change smoothly over time or in block periods
Index q can vary smoothly over time

SS

(stock synthesis)

All of the above… and more

XSA

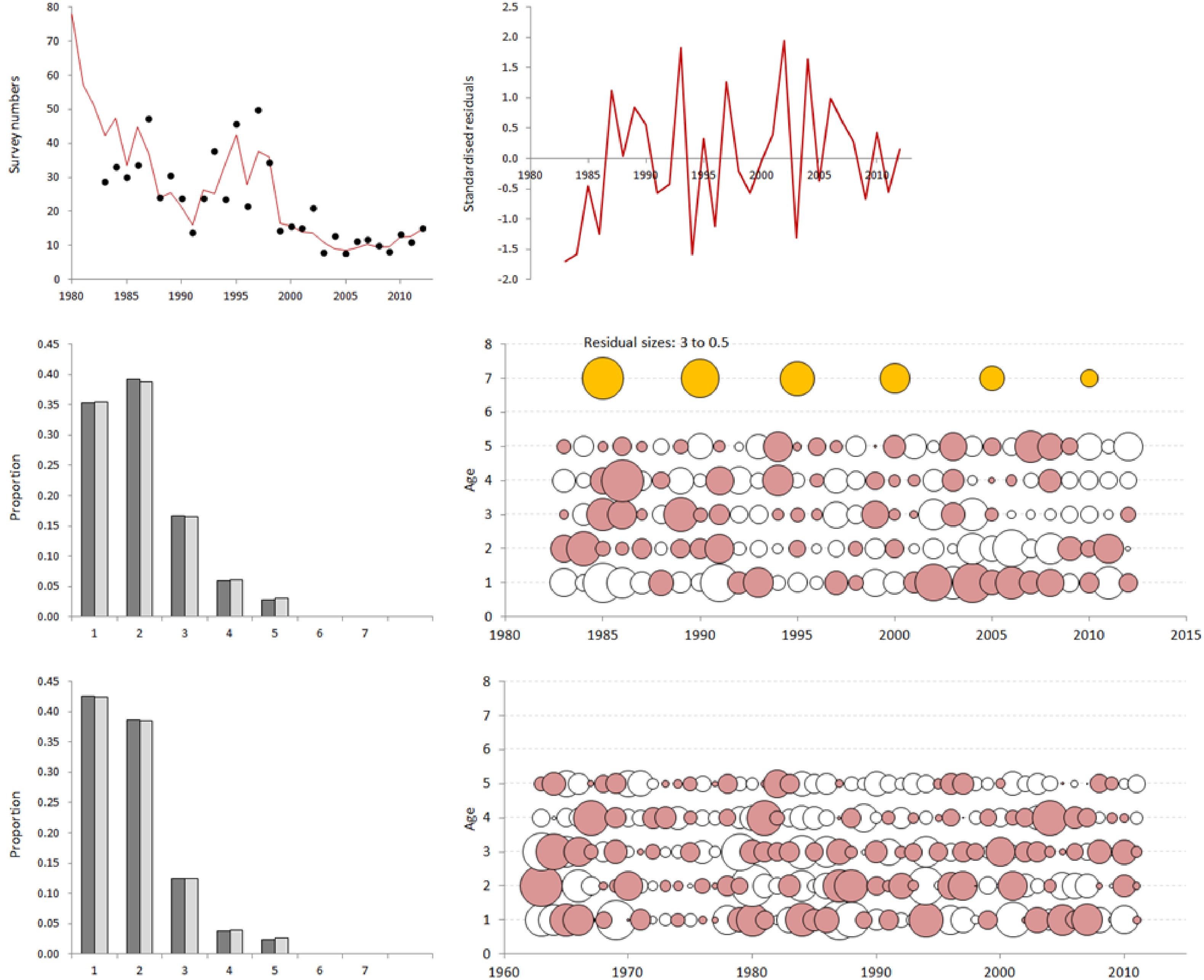


Stoch-ASPM

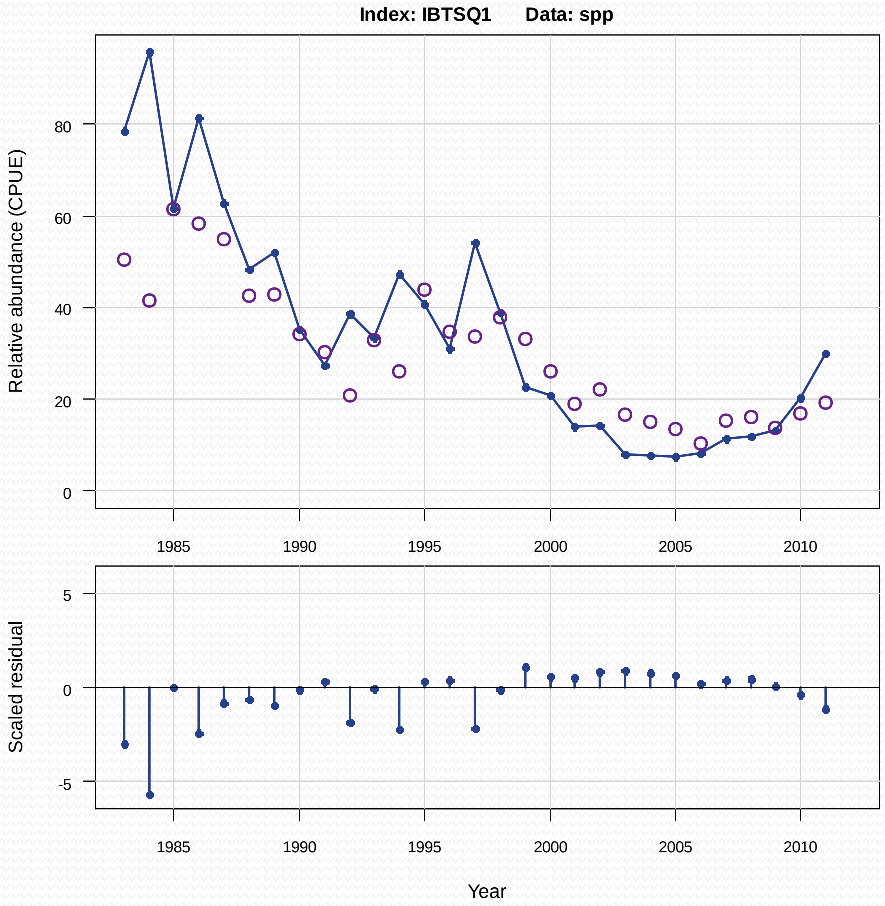
Survey

Survey

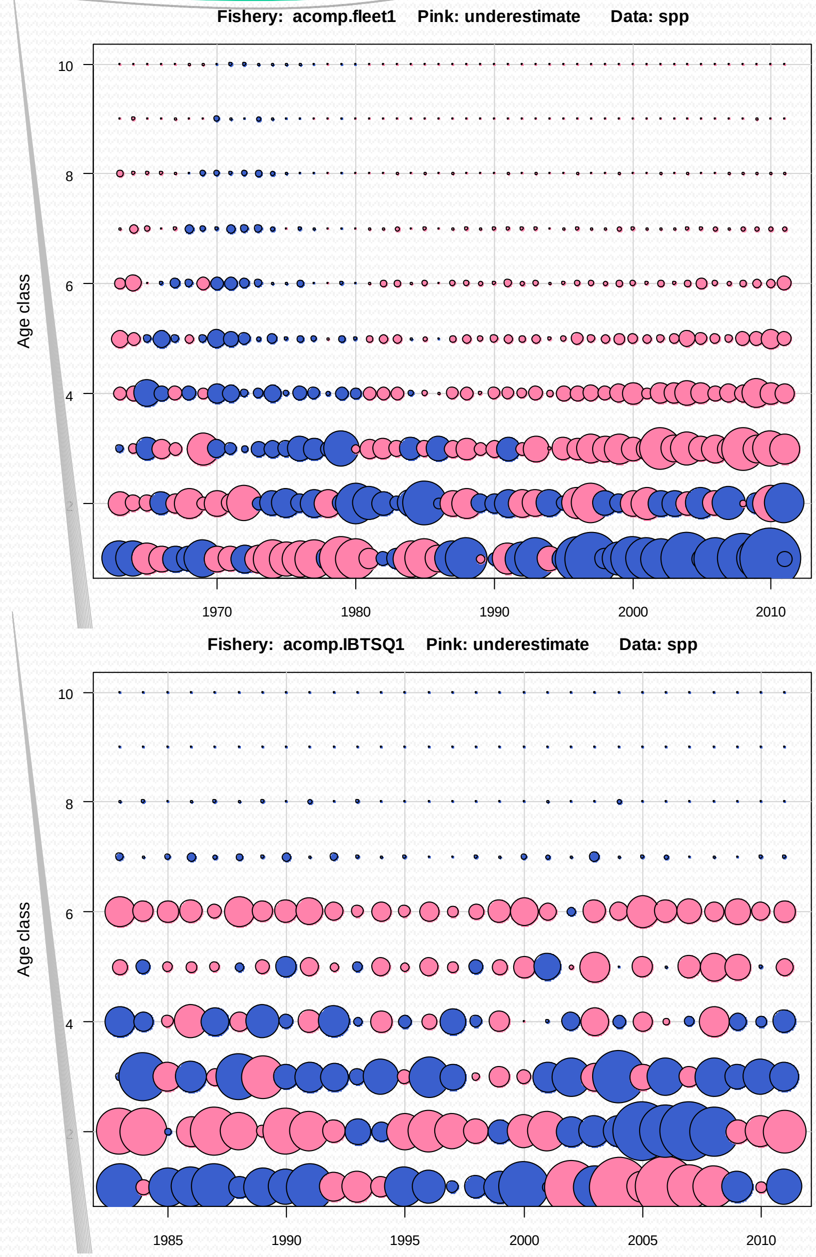
Commercial



BAM

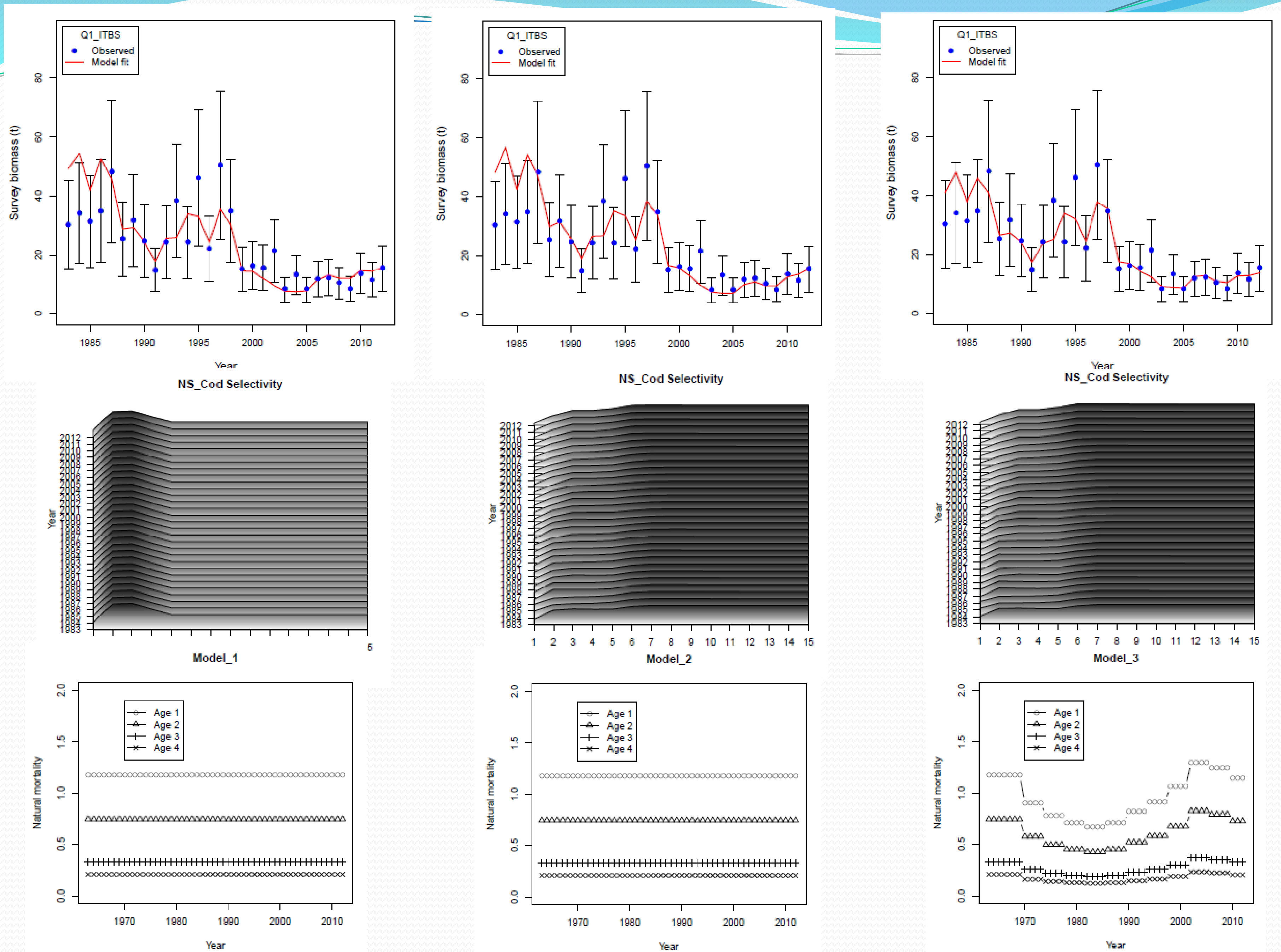


Index: IBTSQ1 Data: spp
Relative abundance (CPUE)
0
20
40
60
80
1985
1990
1995
2000
2005
2010
Scaled residual
5
0
-5
Year

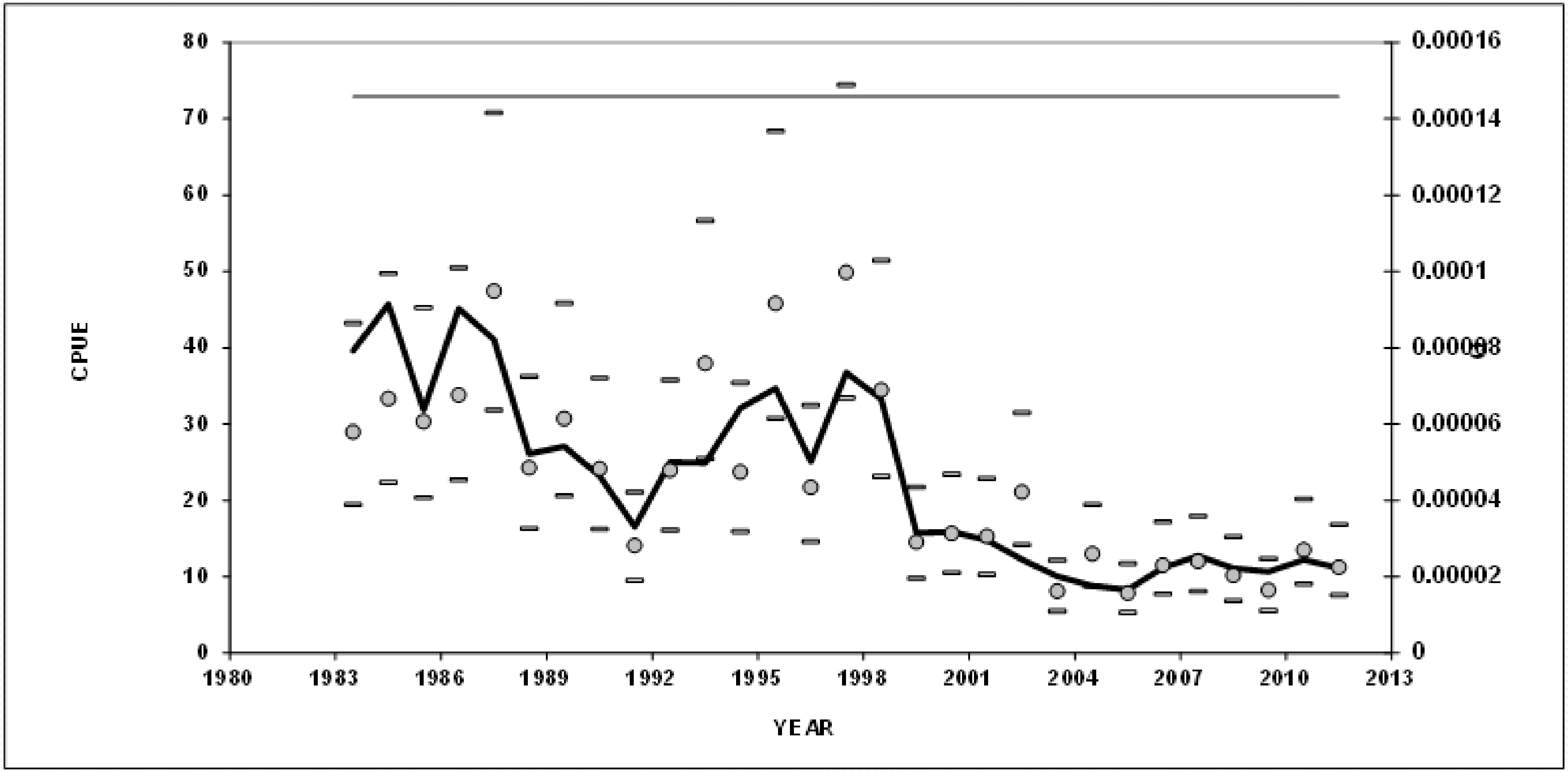


Fishery: acomp.fleet1 Pink: underestimate Data: spp
Age class
1970
1980
1990
2000
2010
Fishery: acomp.IBTSQ1 Pink: underestimate Data: spp
Age class
1985
1990
1995
2000
2005
2010

AMAK: 3 models

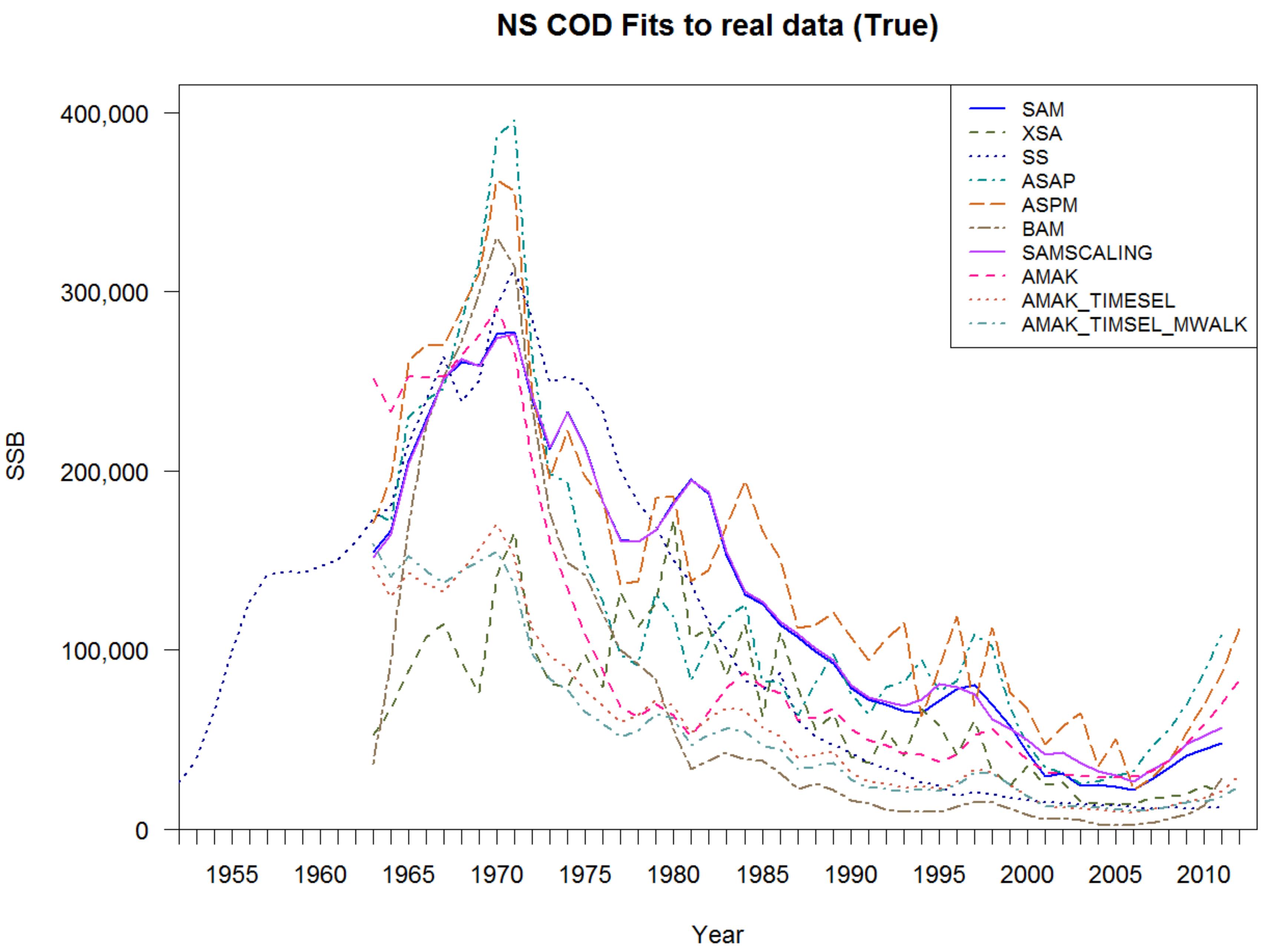


SS

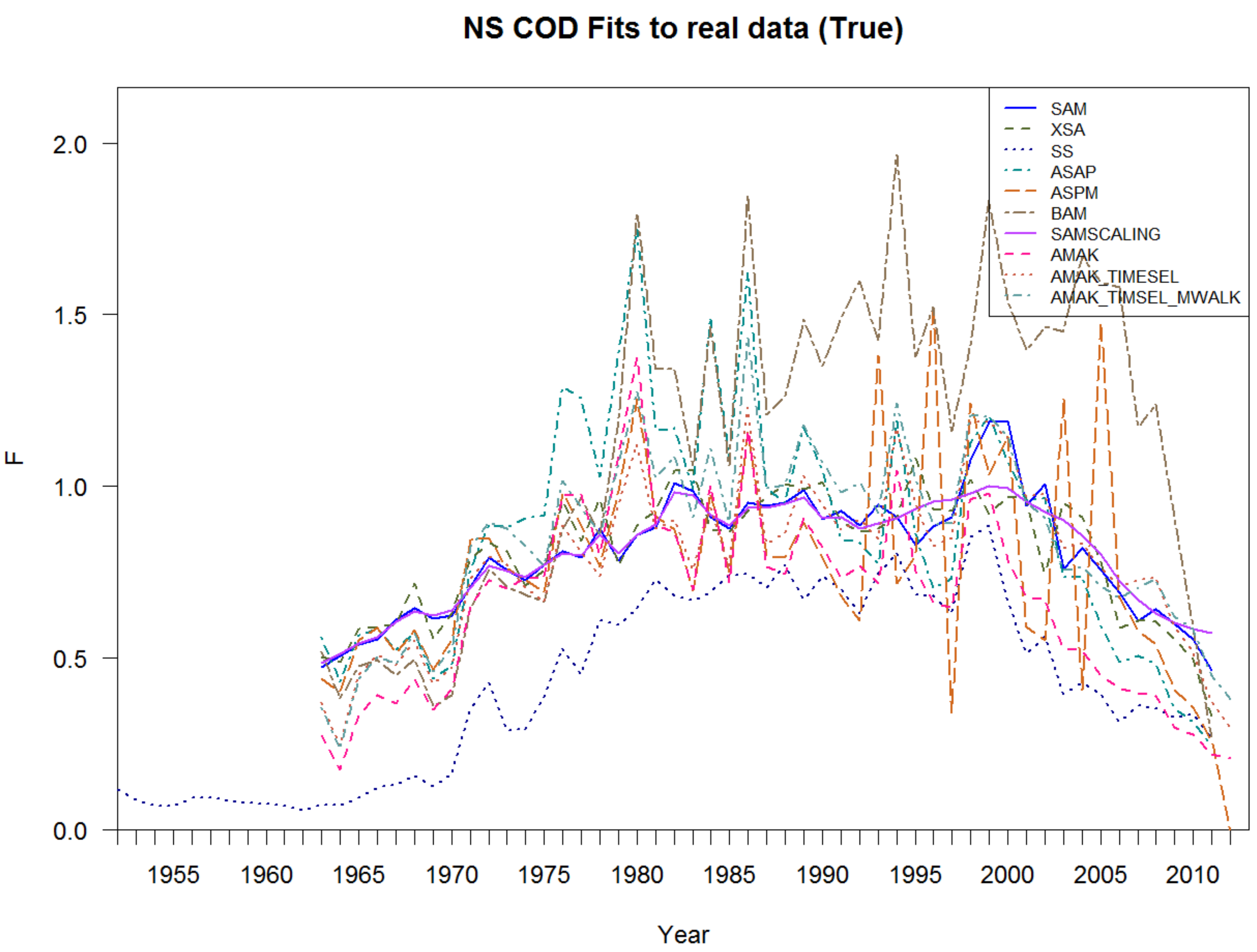


CPUE
YEAR
0
10
20
30
40
50
60
70
80
1980
1983
1986
1989
1992
1995
1998
2001
2004
2007
2010
2013
0
0.00002
0.00004
0.00006
0.00008
0.0001
0.00012
0.00014
0.00016

All models: SSB



All models: F



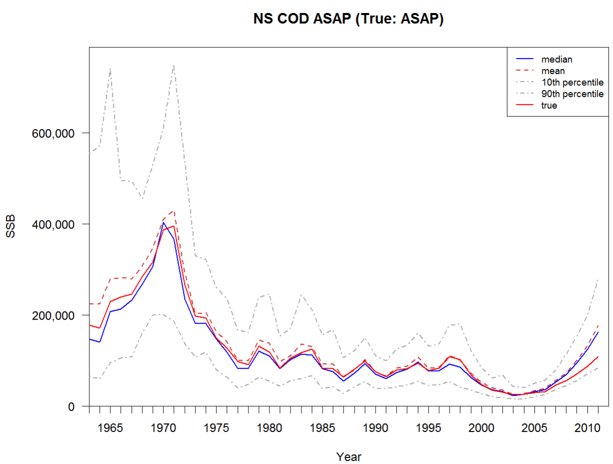
NS COD Fits to real data (True)
SAM
XSA
SS
ASAP
ASPM
BAM
SAMSCALING
AMAK
AMAK_TIMESEL
AMAK_TIMSEL_MWALK
F
2.0
1.5
1.0
0.5
0.0
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
Year

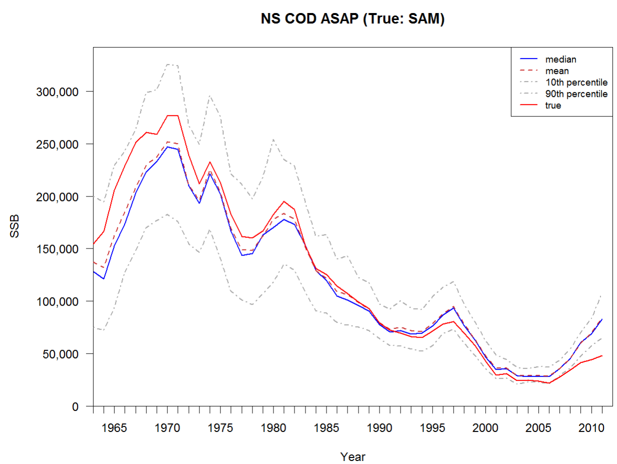
ASAP assessment on alternative operating models: SSB

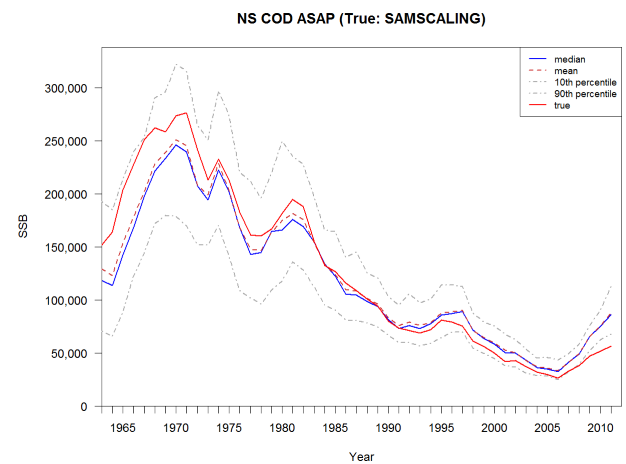
ASAP

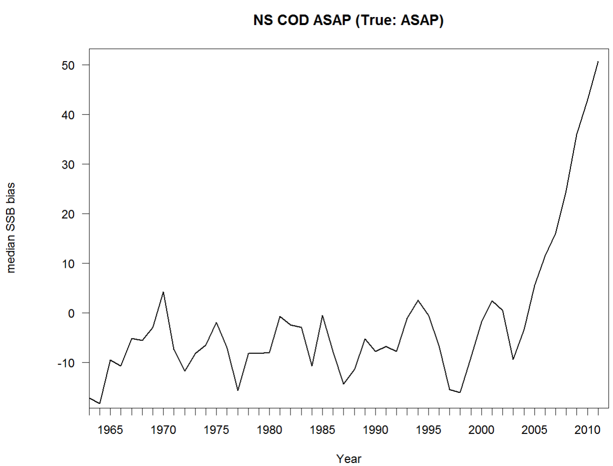
SAM

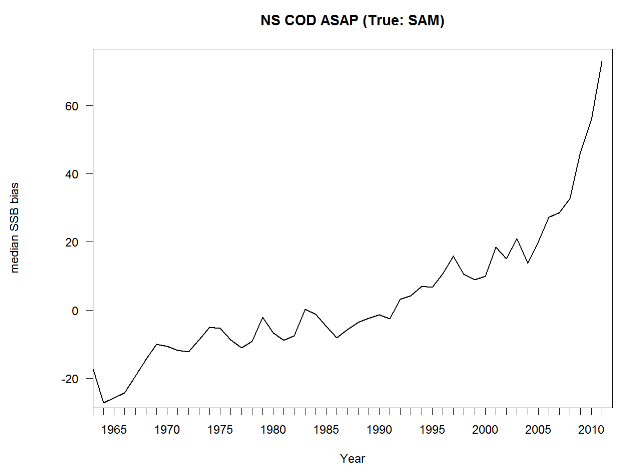
SAM scaling

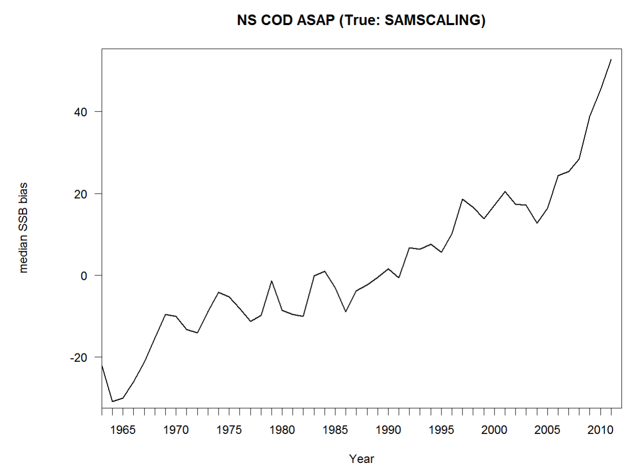












ASAP assessment on alternative operating models: F

ASAP | SAM | SAM scaling

Preliminary Conclusions

<u>Model fits to real data</u>:

If models represent plausible alternatives, then model structure uncertainty is likely being underestimated in MSEs conducted within ICES

<u>Model fits to pseudo data</u>:

The ASAP configured for NS cod has a worrying tendency to overestimate stock size and underestimated fishing mortality, particularly in the most recent years, a crucial period for management decisions

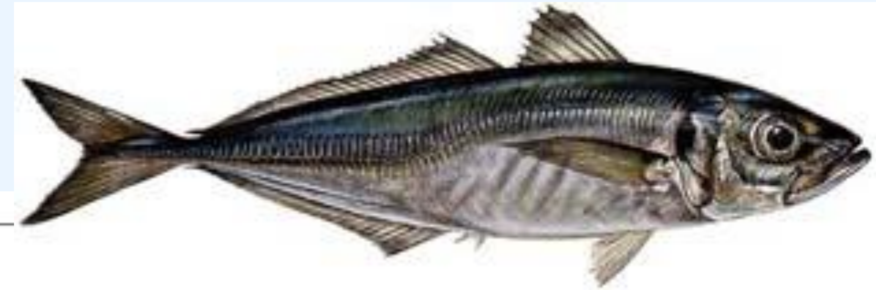
Southern horse mackerel: fits to real and simulated data

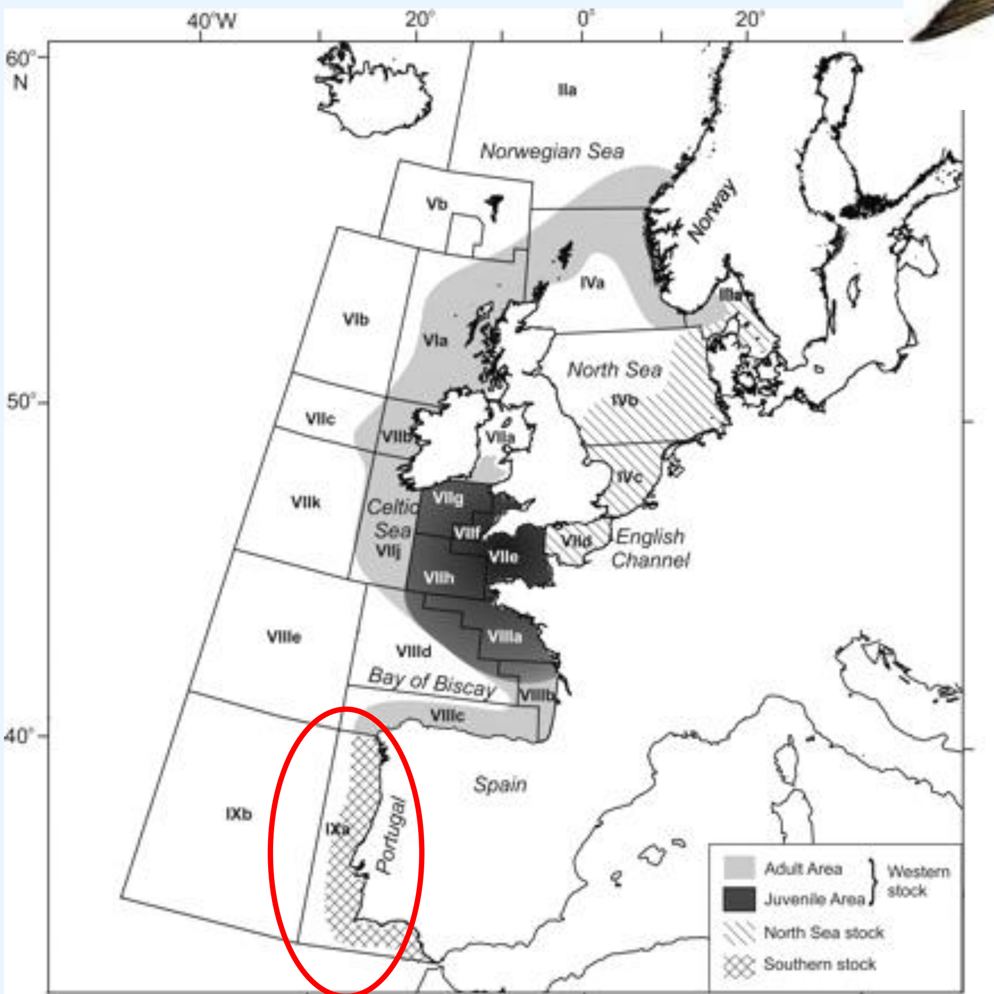
Carmen Fernández, ICES

Based on work from: M. Azevedo (IPMA, PT), N. Brites (IPMA, PT), C. Canales (IFOP, CL), G. Costas (IEO, SP), J. Deroba (NOAA, US), J. Ianelli (NOAA, US), A. Murta (IPMA, PT)

ICES International Council for the Exploration of the Sea
CIEM Conseil International pour l'Exploration de la Mer

SOUTHERN HORSE MACKEREL



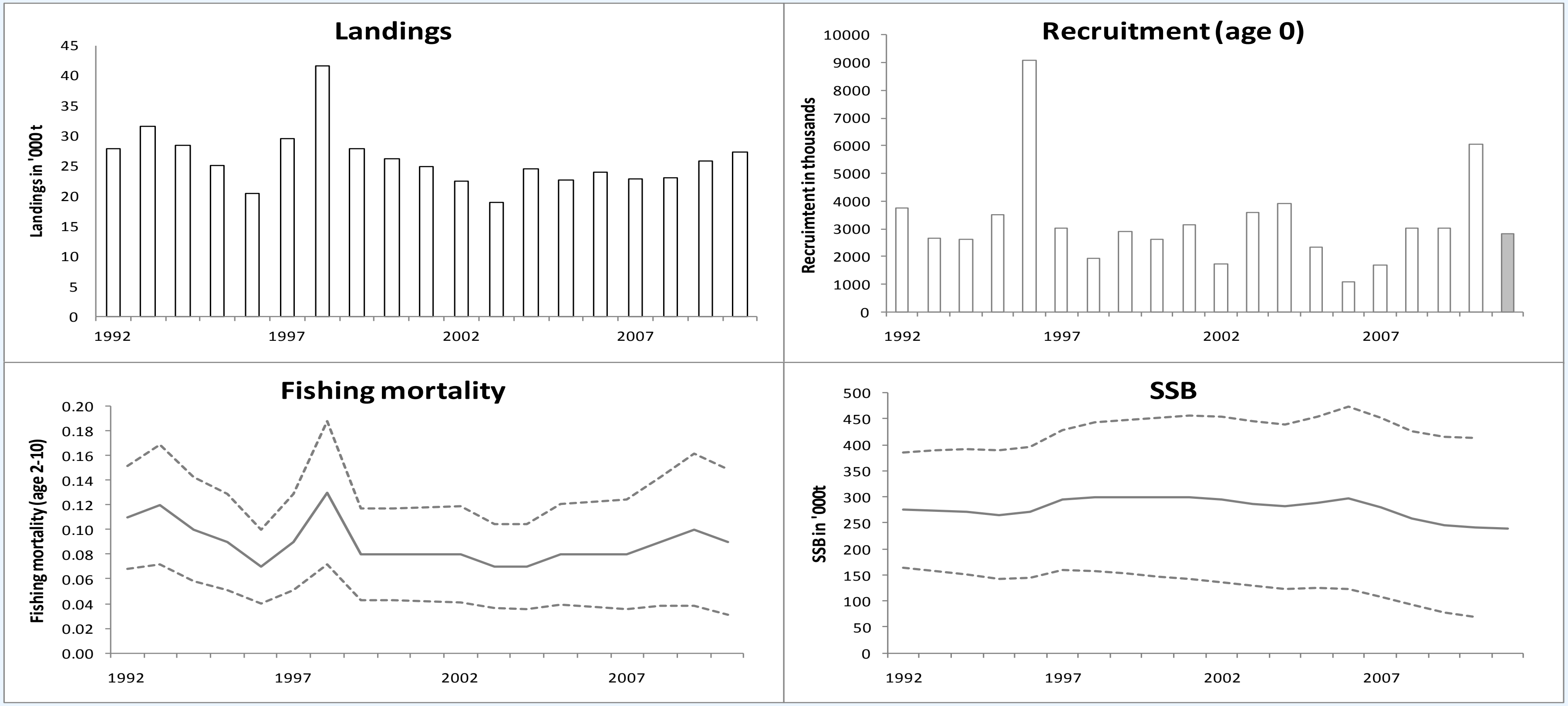


Caught by:

- Bottom trawl
- Purse seine
- Artisanal

ICES conducts assessment and provides advice annually

This study: based on assessment conducted in 2011 (with data until 2010)



- Catches quite stable
- F(ages 2-10) ~ 0.10
- Variable recruitment, with occasional large peaks
- Very wide confidence intervals

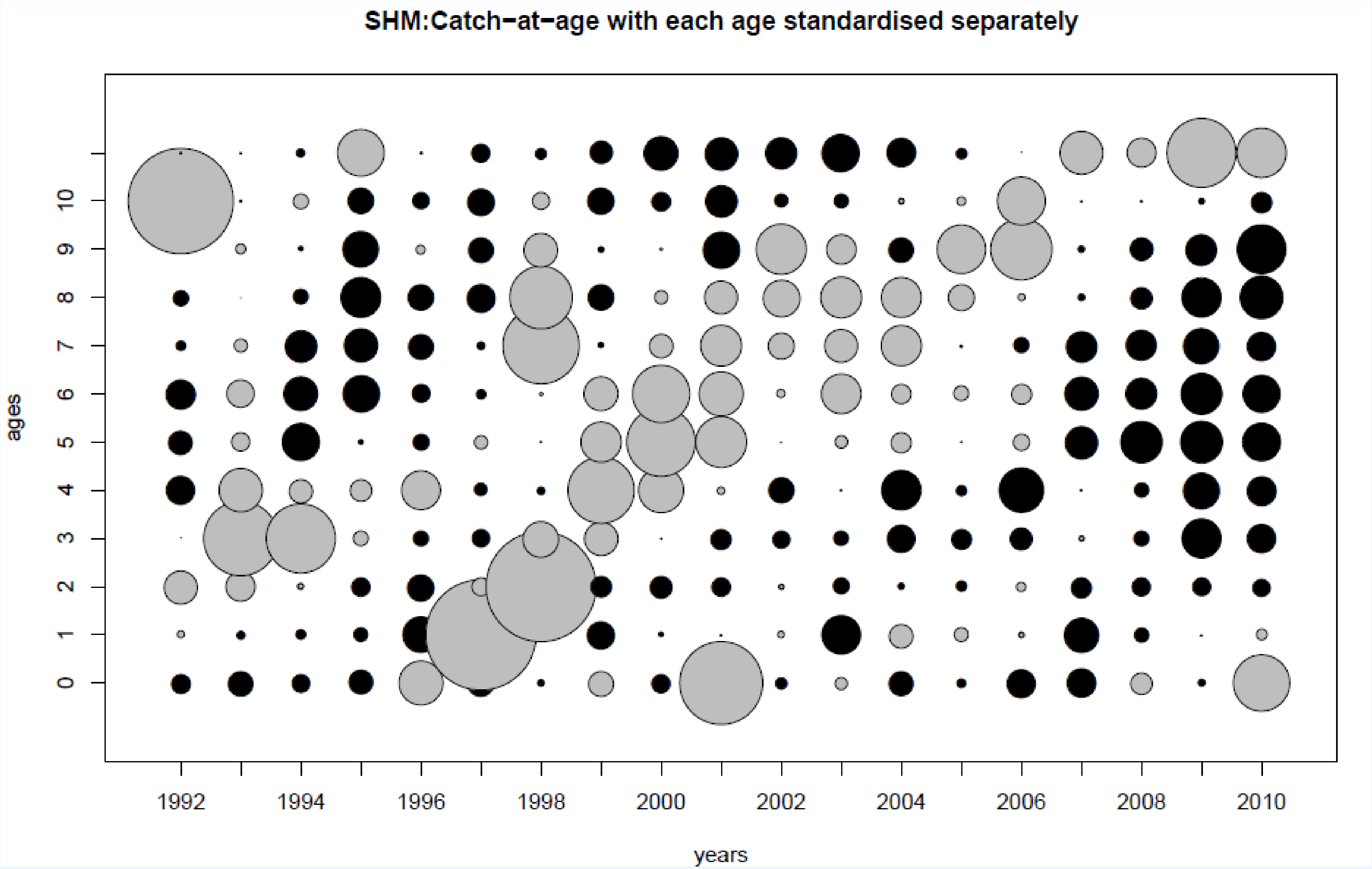
ICES
CIEM

INPUT DATA (years 1992-2010)

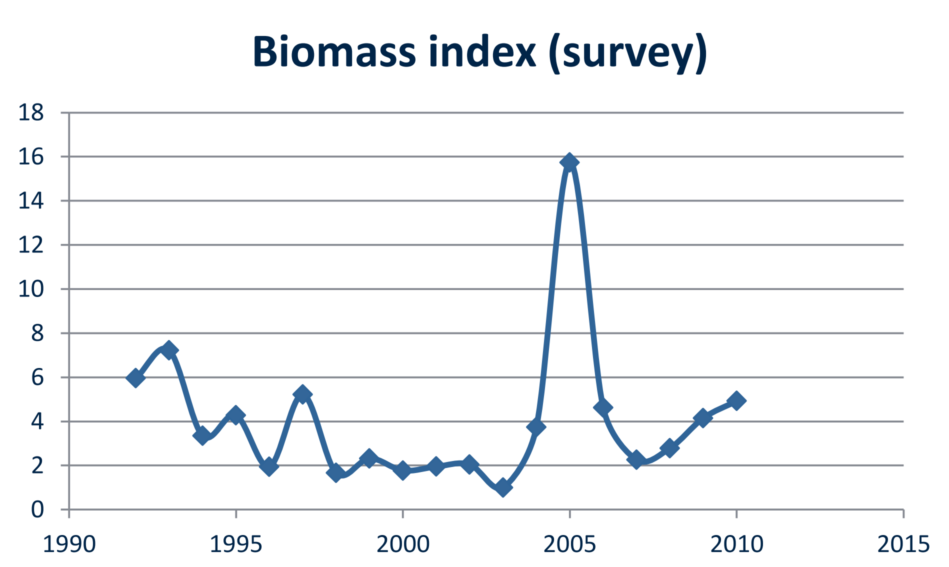
- Annual catch and proportions at age [ages 0-11+]
- Annual abundance index (1 survey) and proportions at age
- Annual mean weight-at-age in catch & stock
- Age structure of data: derived from length-frequency sampling and ALKs
- M (age-dependent, higher for younger ages; constant over years)
- Proportion mature-at-age (constant over the years)

ICES
CIEM

Catch-at-age shows some strong cohorts:



Bubble plot displays, for each age separately: { C(a,y) – Mean[C(a,y)] } / St.Dev.[C(a,y)], with Mean and St Dev taken over the years

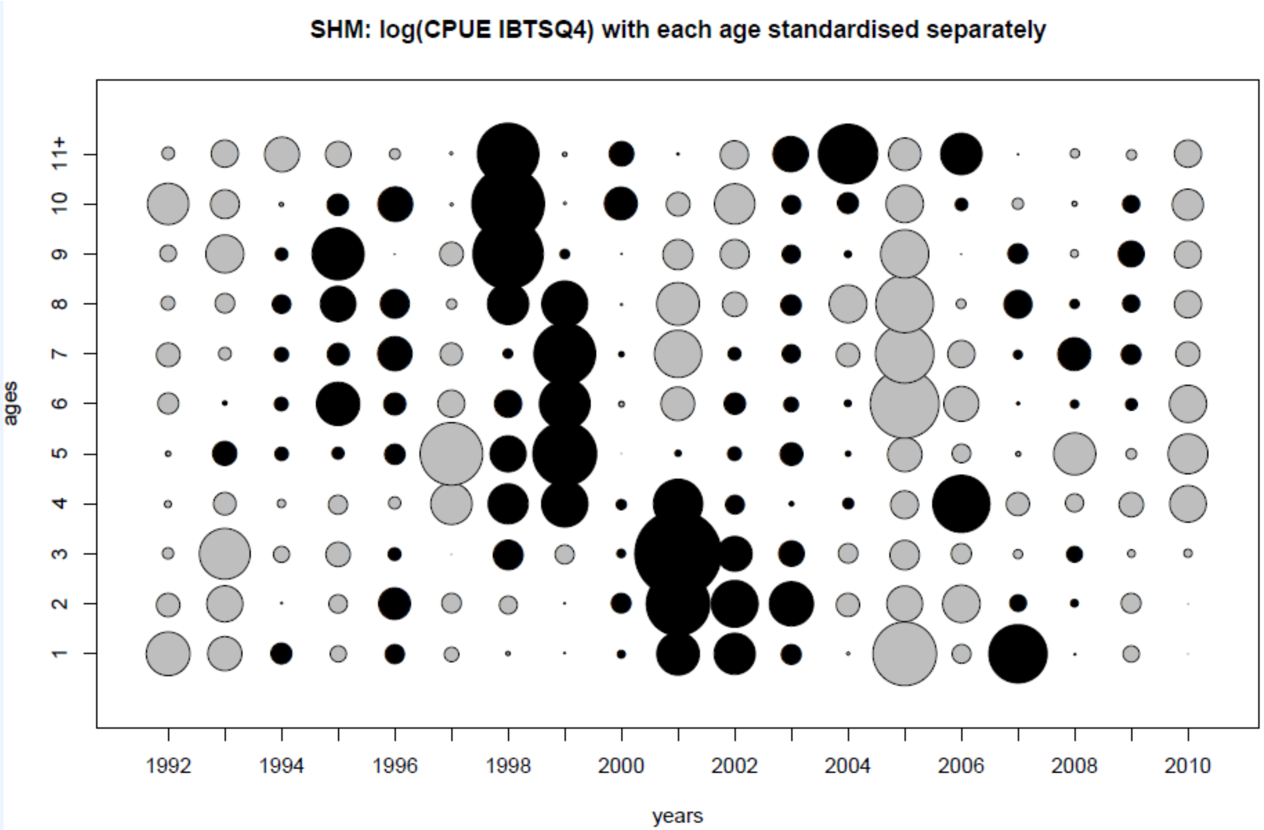


Survey (bottom trawl) shows large variability and year effects

Bubble plot displays, for each age separately:

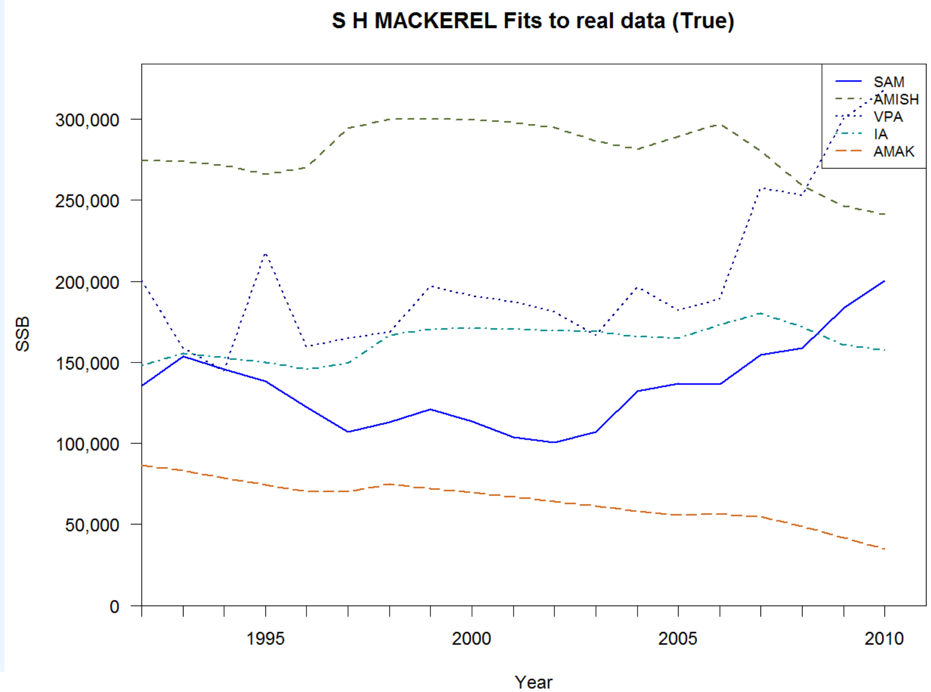
{ln I(a,y) – Mean[ln I(a,y)]} / St.Dev.[ln I(a,y)],

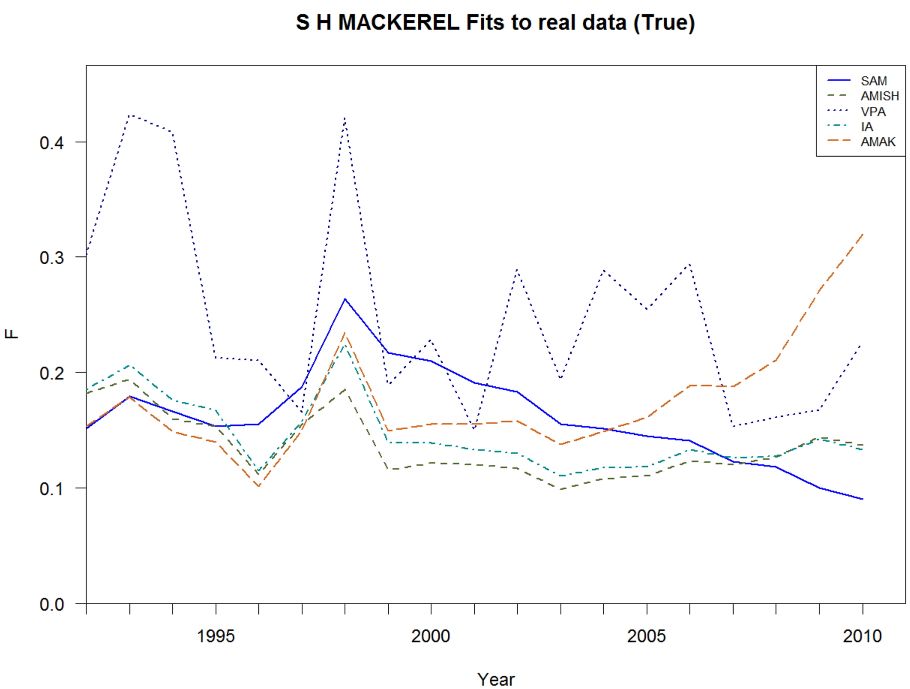
with Mean and St Dev taken over the years



MODELS FITTED:

- AMISH (original fit)
- AMAK
- IA
- SAM
- VPA





SSB of AMAK corresponds only to females

- SSB: differences in overall values (AMISH higher) and trends:
  - ➢ AMISH, IA: stable, some recent decrease
  - ➢ AMAK: decreasing throughout
  - ➢ SAM, VPA: increasing in last decade

F not shown in consistent way for all models

- ➢ AMISH, IA: very similar trends
- ➢ AMAK: increasing
- ➢ SAM: decreasing

ICES CIEM

TRYING TO UNDERSTAND THE DIFFERENCES:

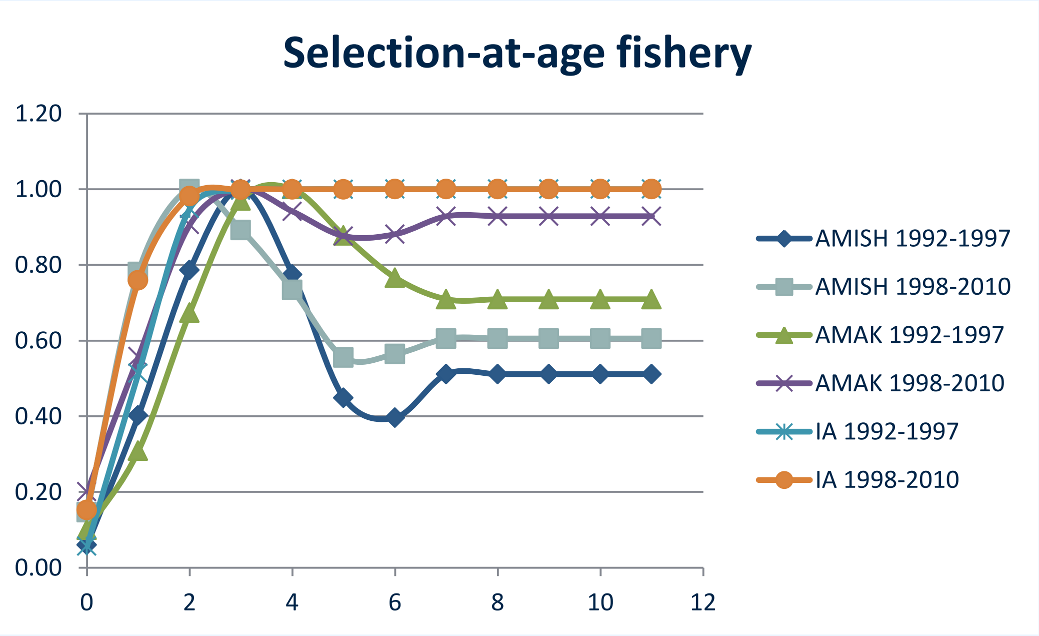
SELECTION-AT-AGE OF FISHERY:

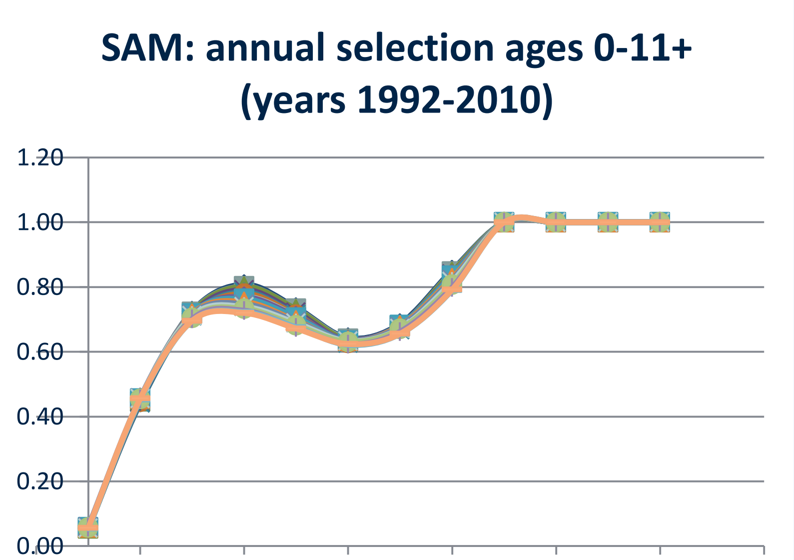
AMISH, AMAK, IA: 2 blocks: 1992-1997, 1998-2010. In each block:

- AMISH, AMAK: a parameter for each age 0-7+, with age-age penalty (penalty stronger in AMAK)
- IA logistic

SAM: ln(F(y,a)) follows (age-correlated) RW in time for each age 0-8+, same variance for all ages

VPA: no modelling, F(oldest age) calculated from preceding ages





SAM estimates selection pattern to be asymptotic

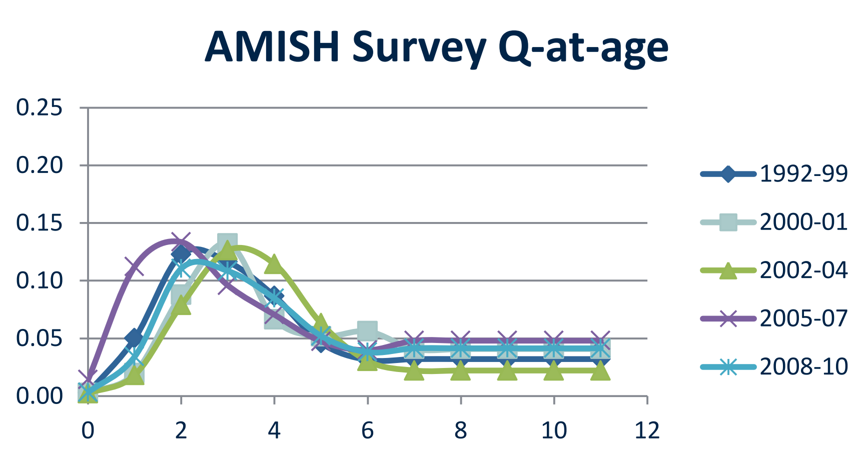
ICES CIEM

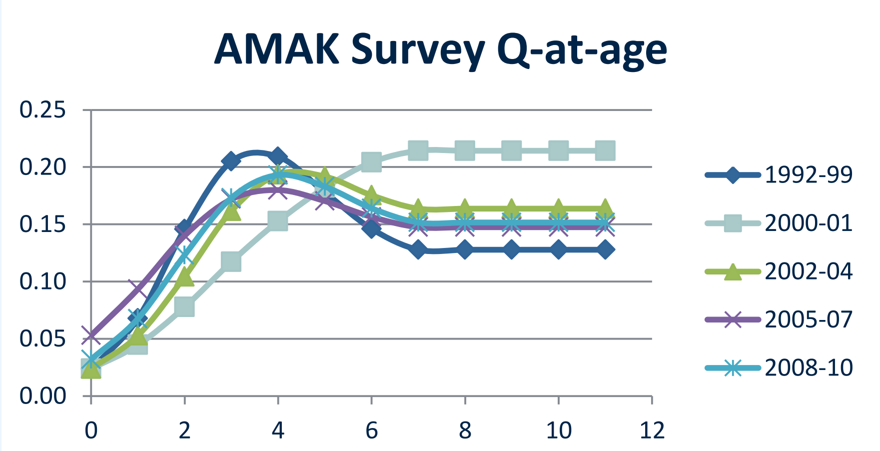
SELECTION-AT-AGE OF SURVEY:

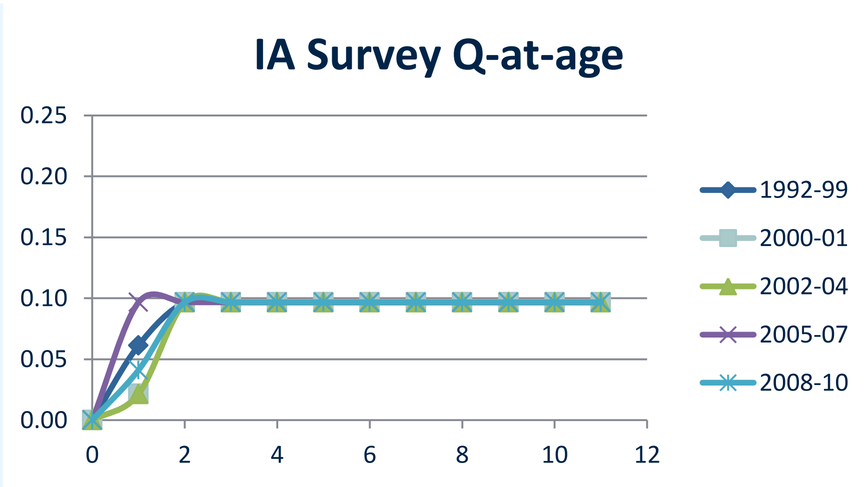
AMISH, AMAK, IA: blocks: 1992-99, 2000-01, 2002-04, 2005-07, 2008-10:

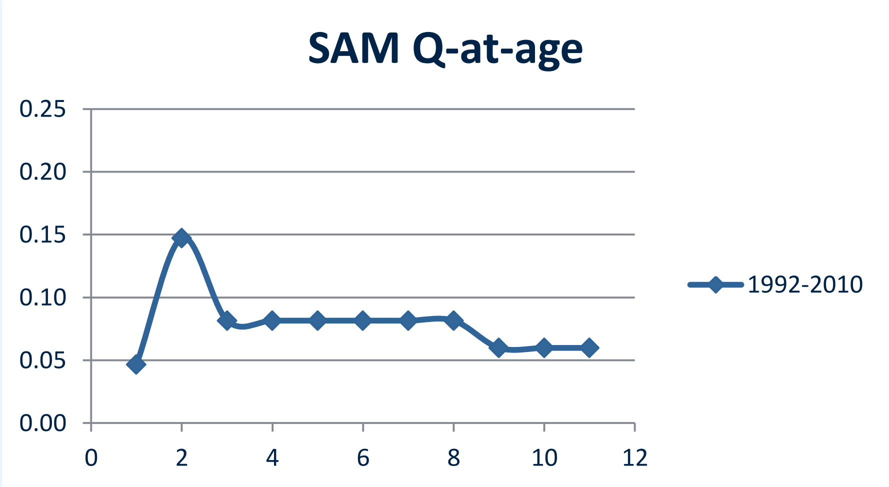
- AMISH, AMAK: a parameter for each age 0-7+, with age-age penalty (penalty stronger in AMAK)
- IA logistic

SAM: catchability for groups of ages, constant over time



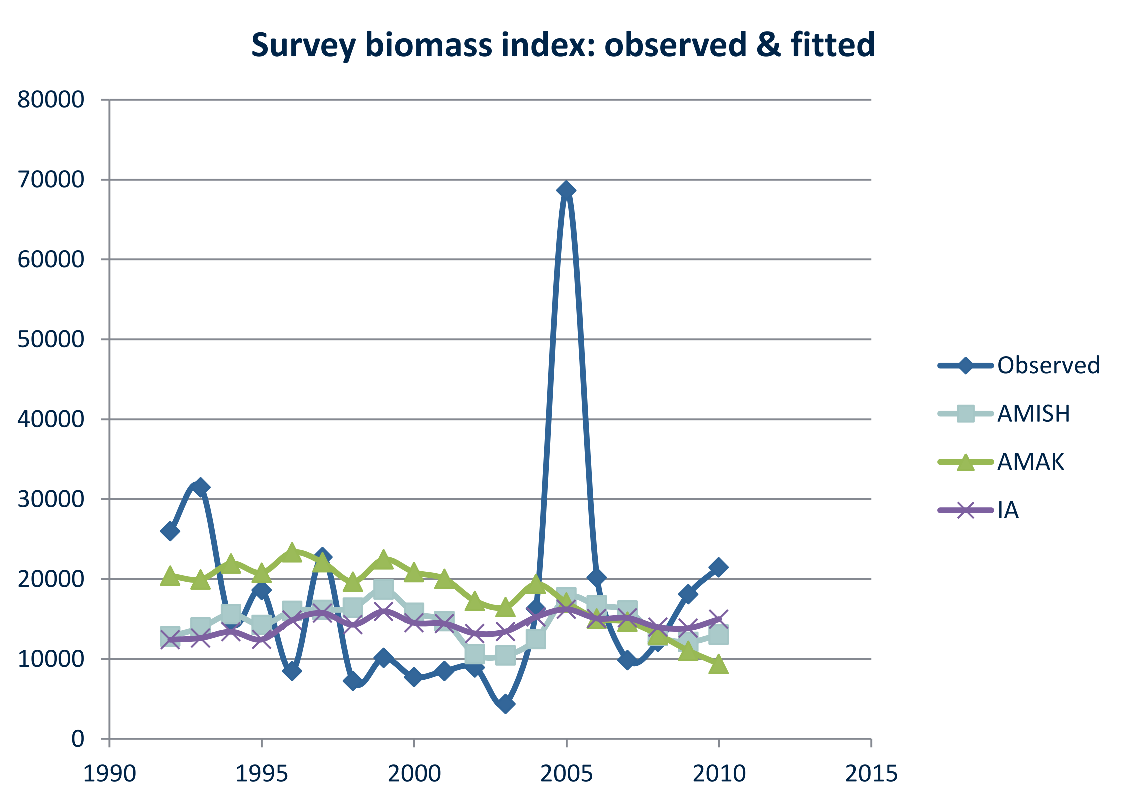




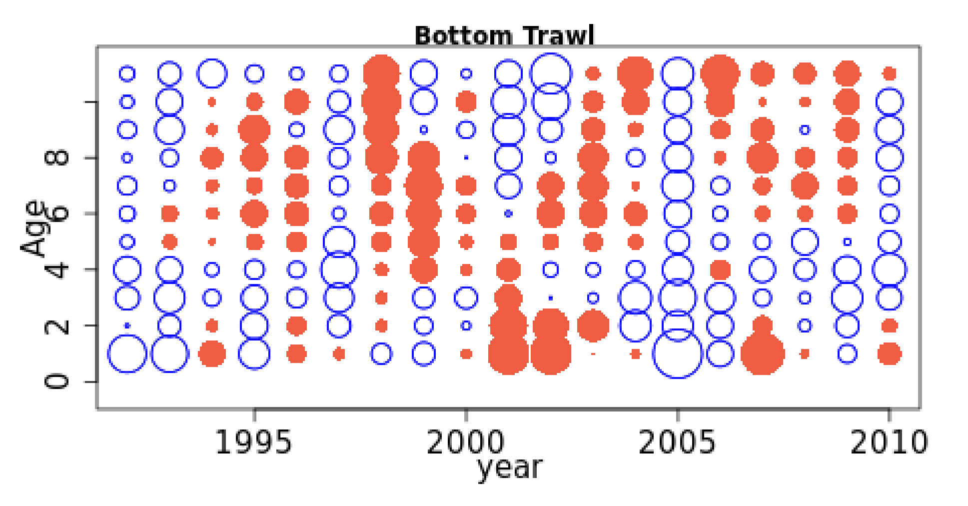


ICES
CIEM

FITS TO SURVEY INDEX



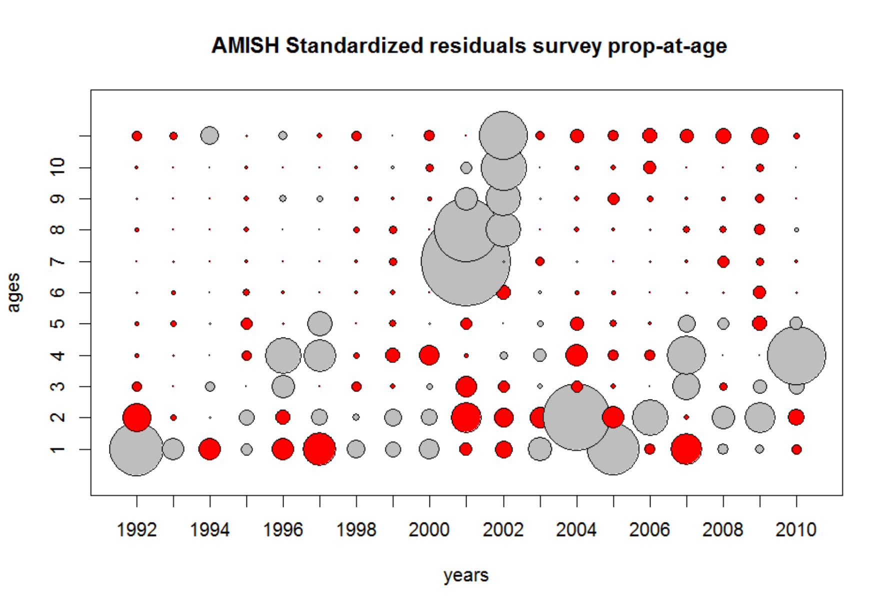
SAM: normalized residuals (red < 0)

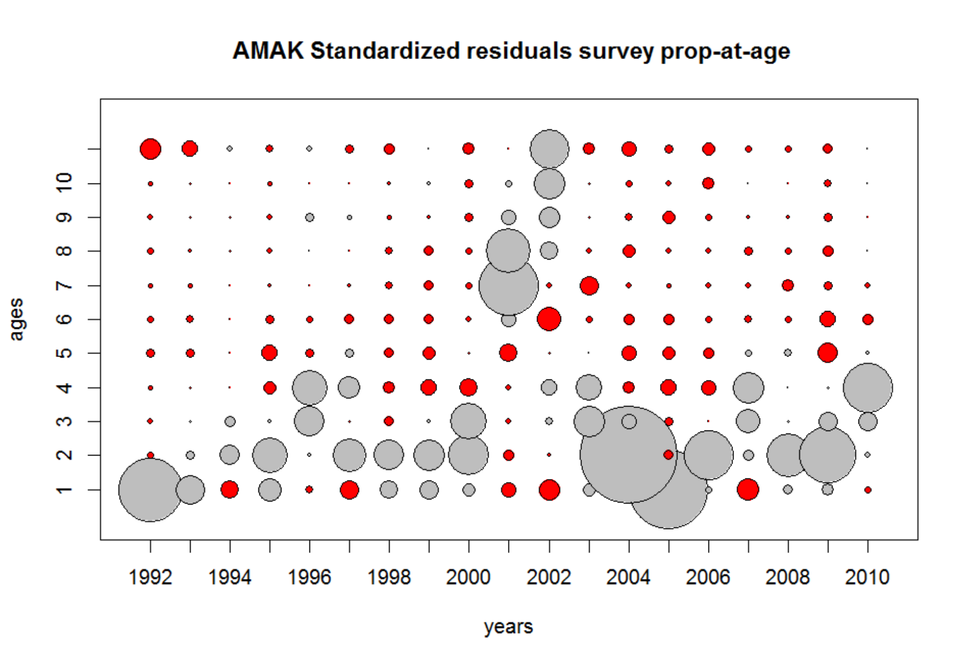


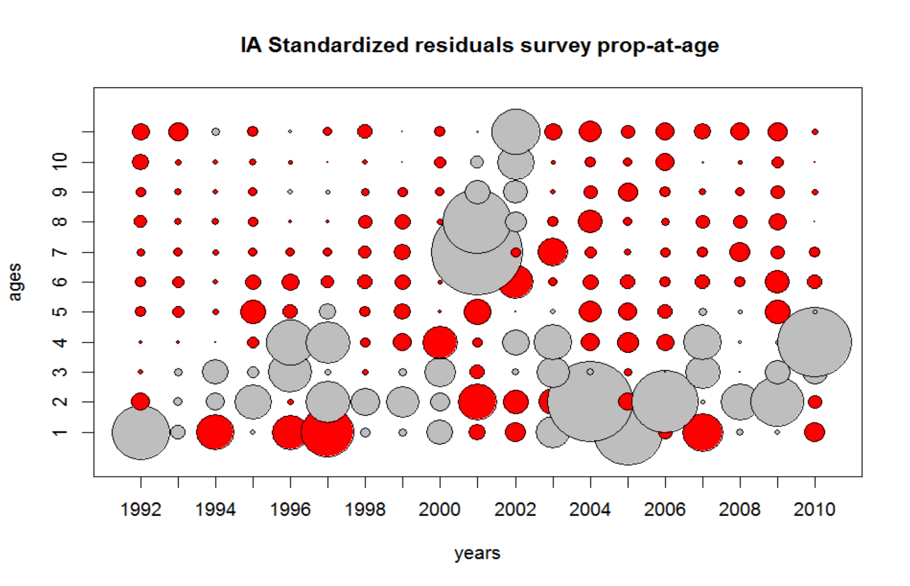
Fits of AMISH, AMAK, IA to survey index

ICES CIEM

RESIDUALS FROM SURVEY PROPORTIONS-AT-AGE:





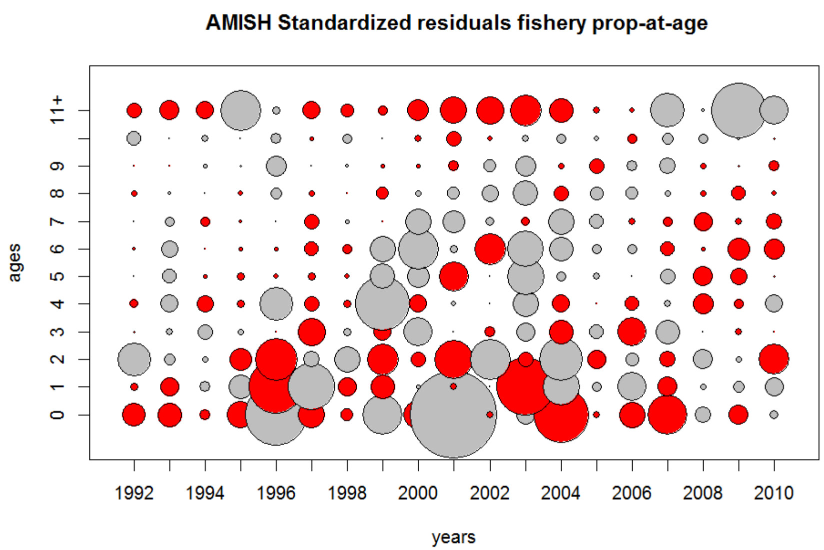


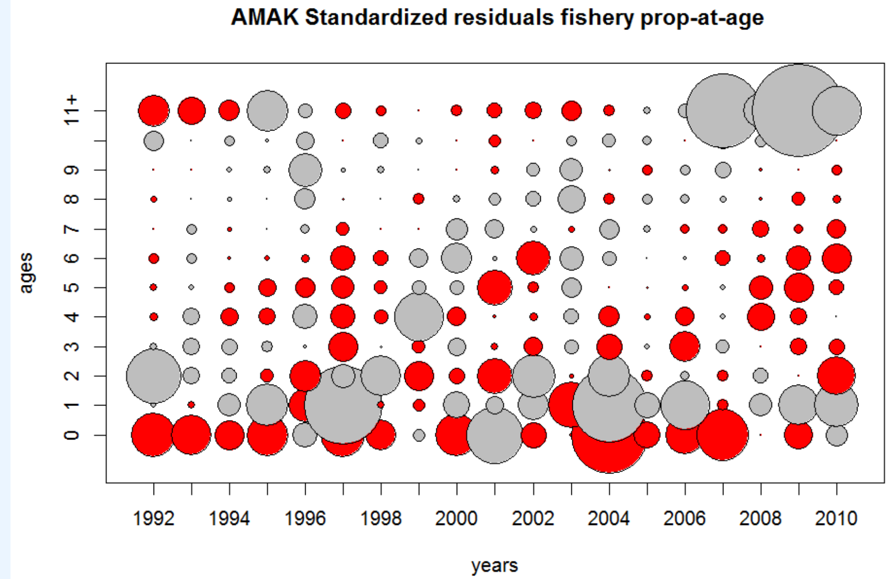
Residual =

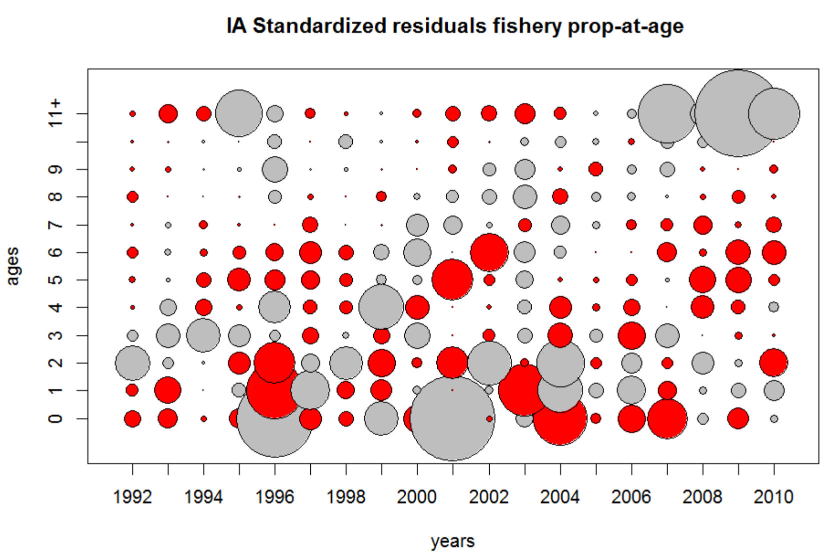
(pobs – pfit) / SQRT{ pfit (1-pfit) }

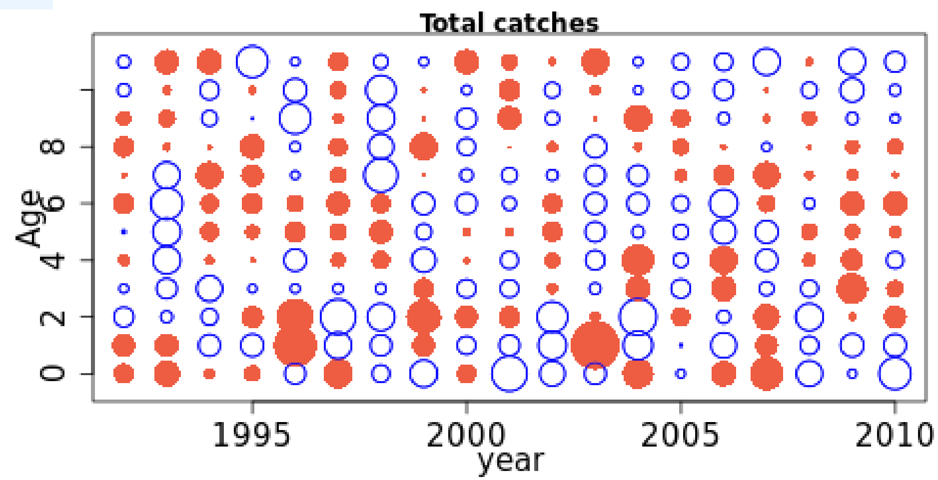
red < 0

RESIDUALS FROM FISHERY:









AMISH, AMAK, IA: residual = (pobs – pfit)/SQRT{ pfit (1-pfit) }; red < 0
SAM: normalized residuals of catch numbers-at-age

CONCLUSIONS FROM FITS TO REAL DATA:

- Substantial differences in SSB and F estimates: overall values and trends

- Different models estimate rather different selection patterns for fishery
- Same comment for survey

➔ Highlights the sensitivity of assessment results to model configuration

- Other differences in models likely also contributing to differences in results (e.g. weight given to different sources of information: CVs, sample sizes, ...)

- Detailed analysis would be necessary to gain further understanding

- Diagnostics (fits, residuals, retrospective analyses,...) and details of each model configuration should be examined carefully

GENERATION OF SIMULATED DATA AND FITS:

Data generated with POPSIM to be consistent with IA fit to real data

"True" annual F-at-age, N-at-age in year one, annual recruitment: IA estimates from fit to real data
➔ N-at-age each year calculated deterministically

Survey selectivity and catchability = IA fit to real data

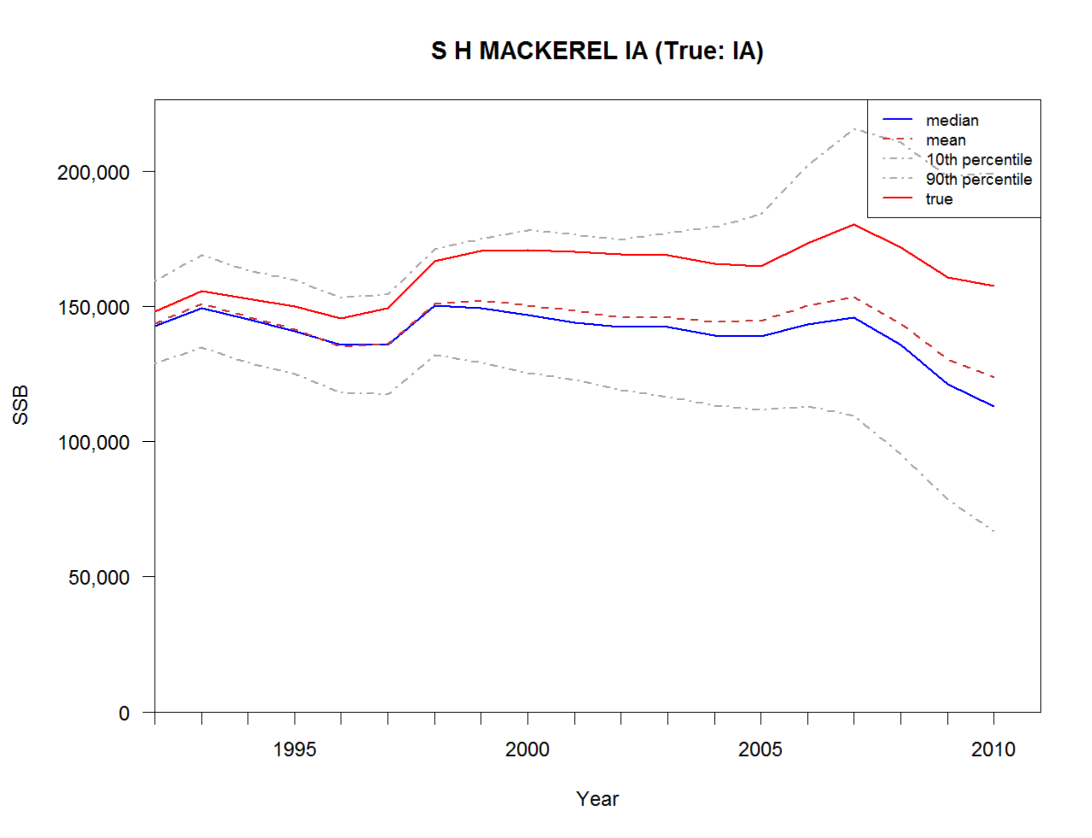
100 datasets generated with POPSIM:

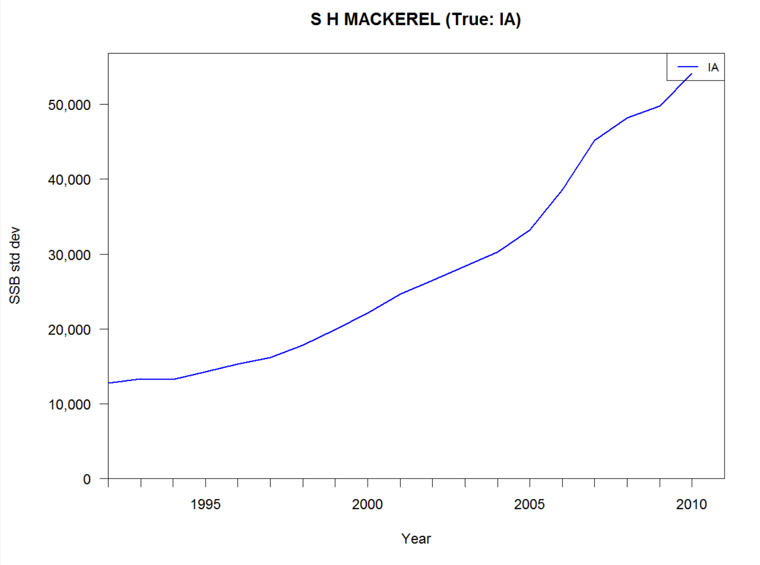
- Catch (weight) ~ Log-Normal with CV=0.4%
- Catch age composition ~ Multinomial with N=100
- Biomass index ~ Log-Normal with CV=70%
- Survey age composition ~ Multinomial with N=20

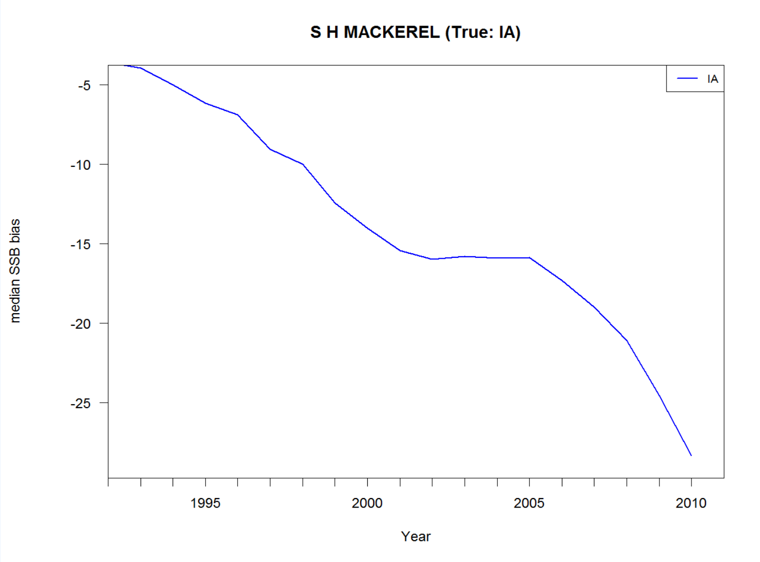
The 100 datasets generated were fit with IA, with same settings as for real data

ICES CIEM

FITS TO SIMULATED DATA: SSB

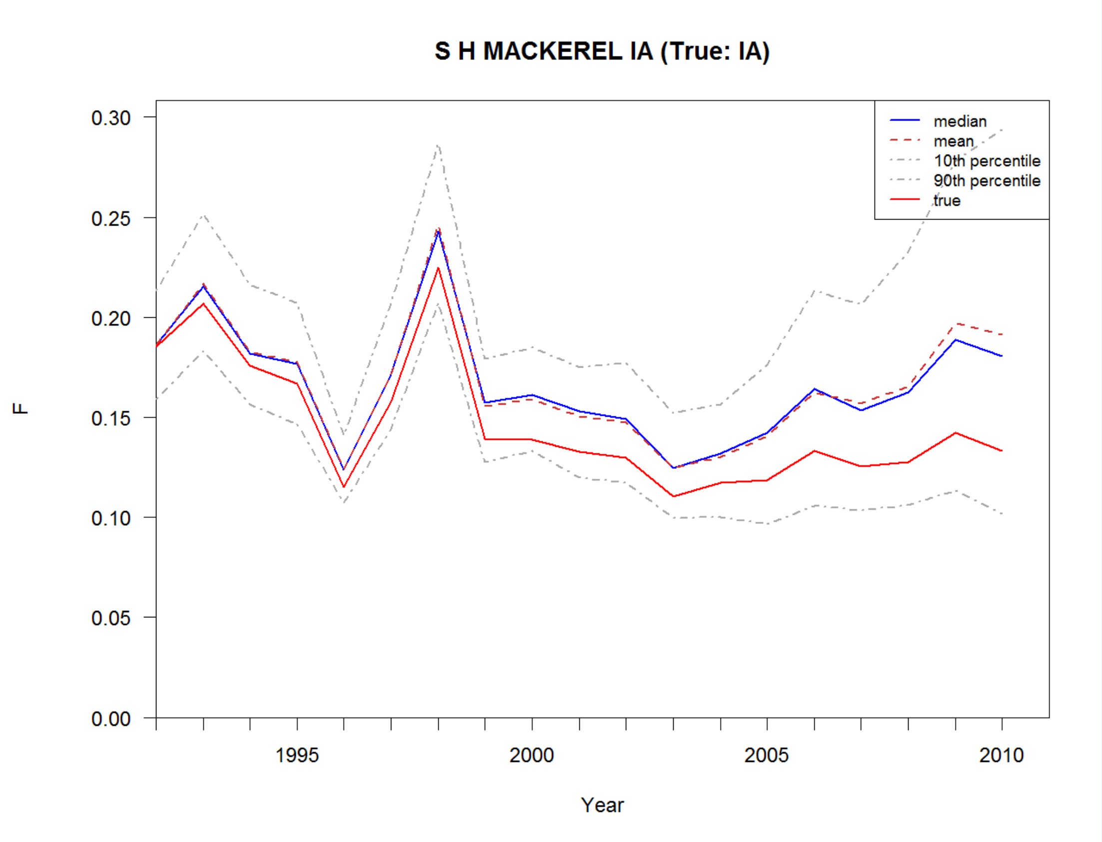


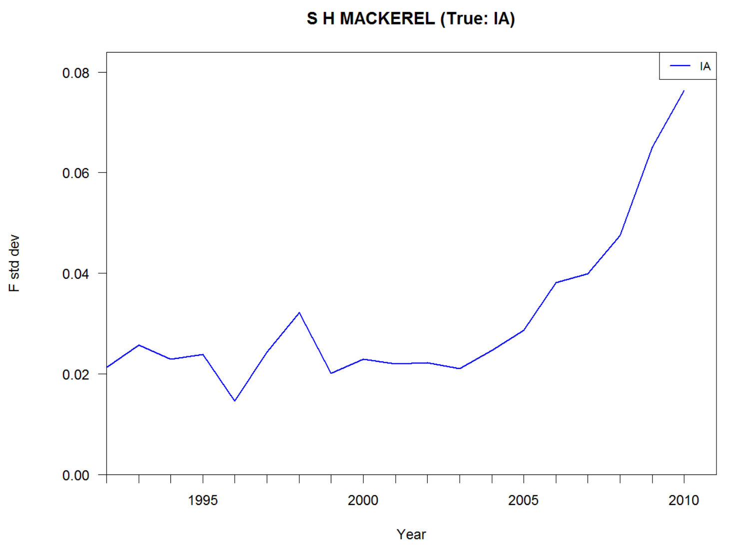


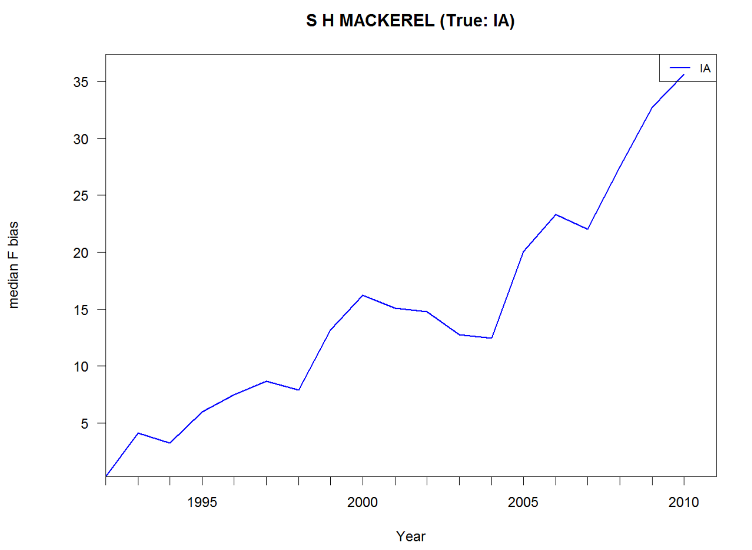


- True SSB inside 80% confidence intervals
- Bias: mean and median of estimates below true SSB
- In recent years: confidence intervals wider and bias larger

FITS TO SIMULATED DATA: F







- True F inside 80% confidence intervals
- Bias: mean and median of estimates above true F
- In recent years: confidence intervals wider and bias larger

ICES CIEM

FITS TO SIMULATED DATA:

- Not clear what causes the bias
- Most likely cause is some some systematic departure in the way simulated data have been generated (with POPSIM) with respect to assumptions made in IA model fits (e.g. weights-at-age or, possibly, differences in bias correction assumptions)
- Understanding this would require further exploration

ICES CIEM

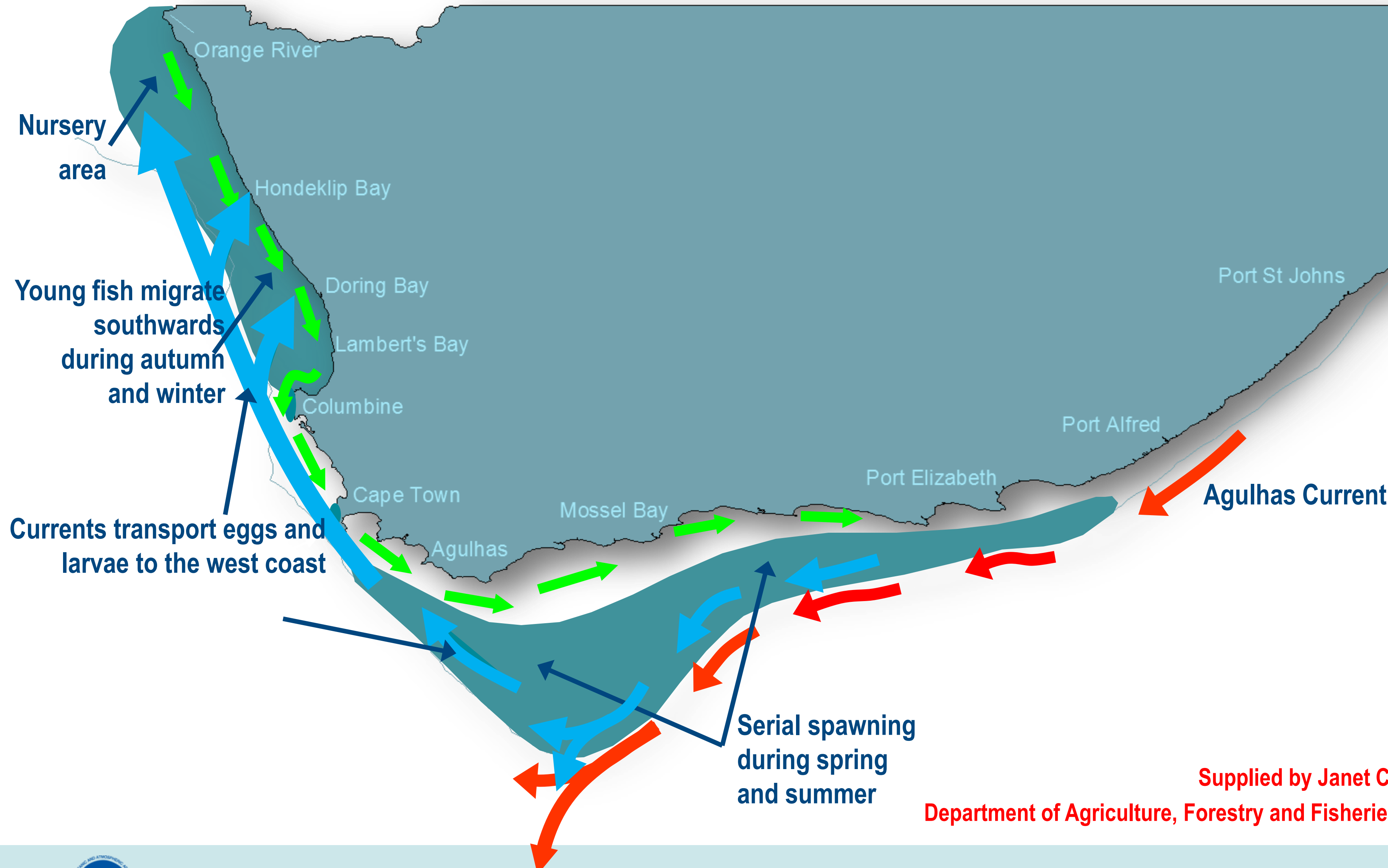
Thank you

SISAM case study using real data and pseudo-data: South African anchovy

Grant Thompson
Alaska Fisheries Science Center

With contributions by Carryn de Moor
Marine Resource Assessment and Management Group
University of Cape Town

Biology and annual recruitment cycle

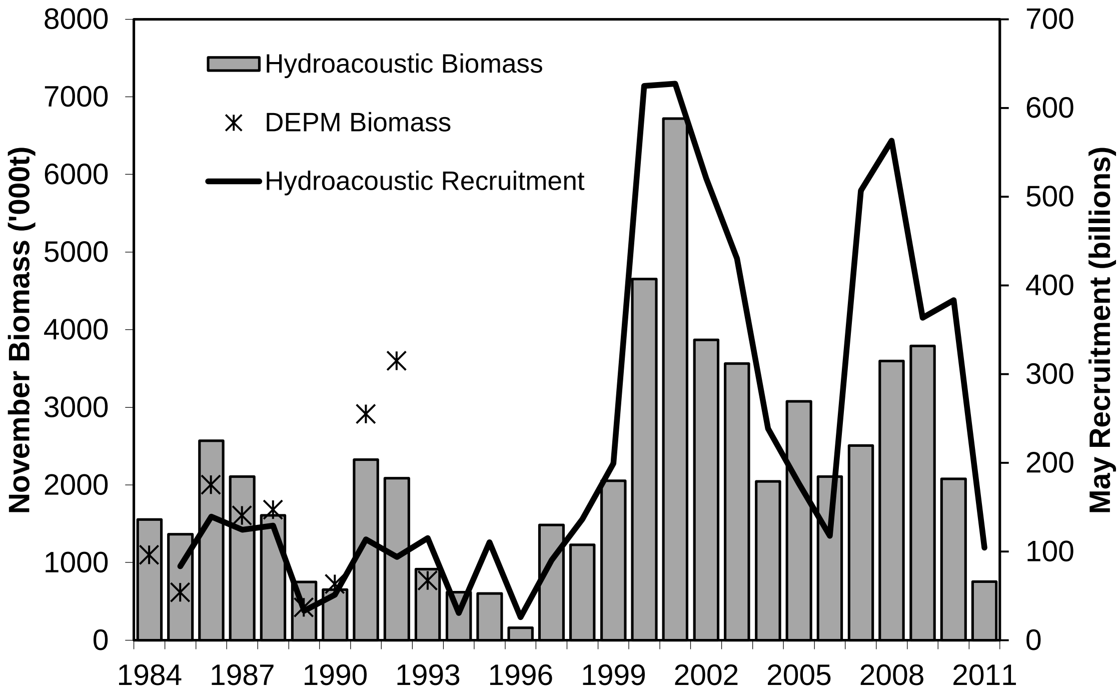


Supplied by Janet Coetzee
Department of Agriculture, Forestry and Fisheries, RSA

Data (1 of 2)

- May hydroacoustic survey: recruitment
- November hydroacoustic survey: SSB
- November DEPM estimates of absolute SSB (1984-1993)

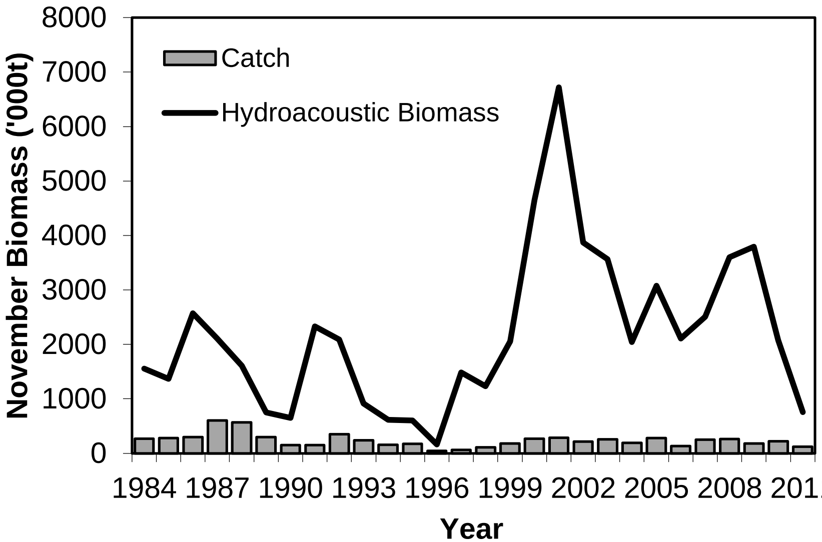
relative abundance



Data (2 of 2)

- Catch at ages 0,1 (length frequencies separated on monthly basis, cut-off lengths differ each year)
- Proportion at age 1 in November survey
- Other annual data:

- timing of May survey
- catch prior to May survey
- November weight-at-age



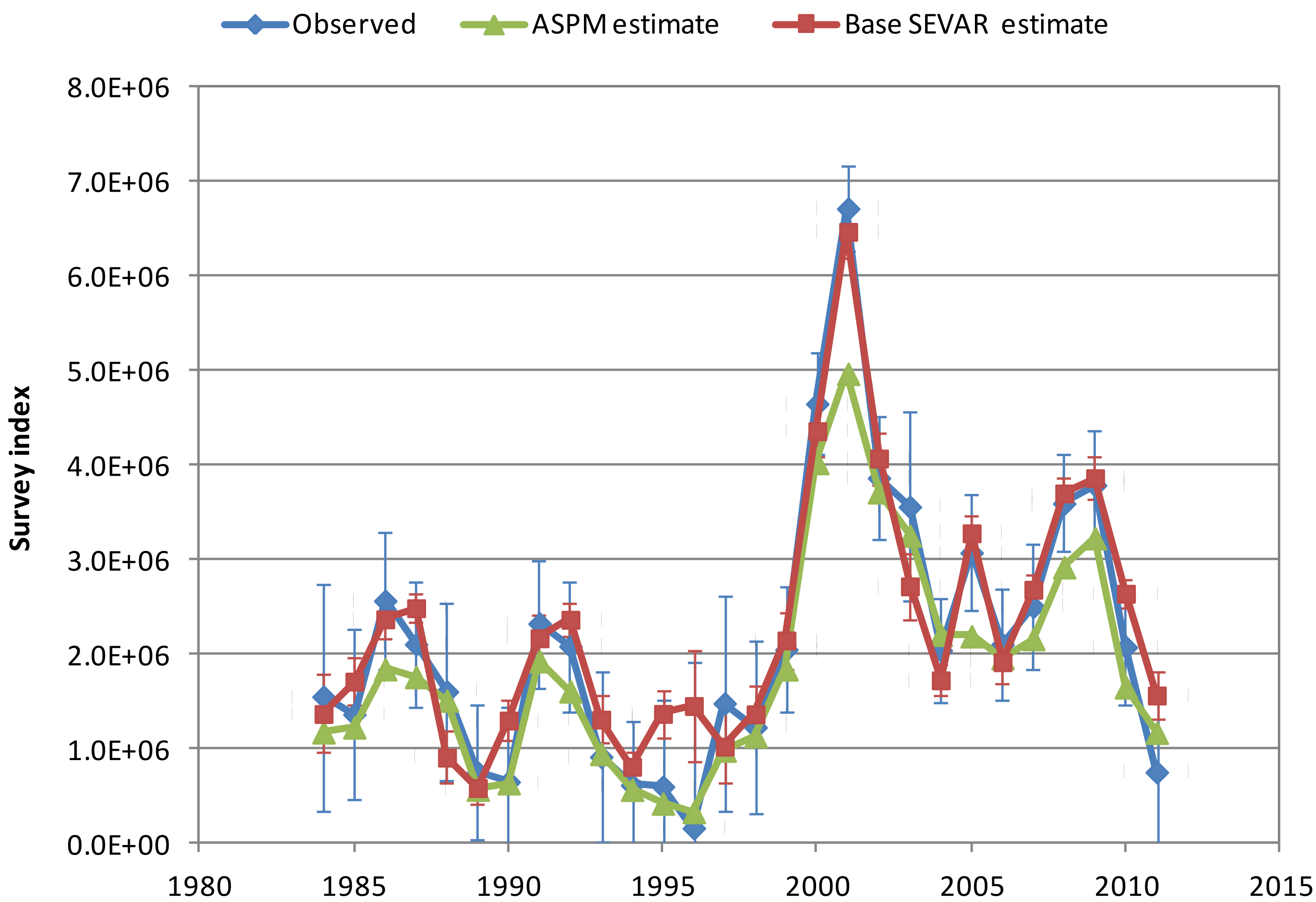
First of two models: ASPM

- Stands for "age-structured production model"
  - This is the model actually used for management, and to which the "real" data file corresponds
  - Serves as operating model, used to project forward when simulation testing management procedures
- Age structured (0 to 4+)
- Annual time steps (November to October)
- Uses all data provided in the SISAM data file
- Not used to generate or fit pseudo-data for SISAM

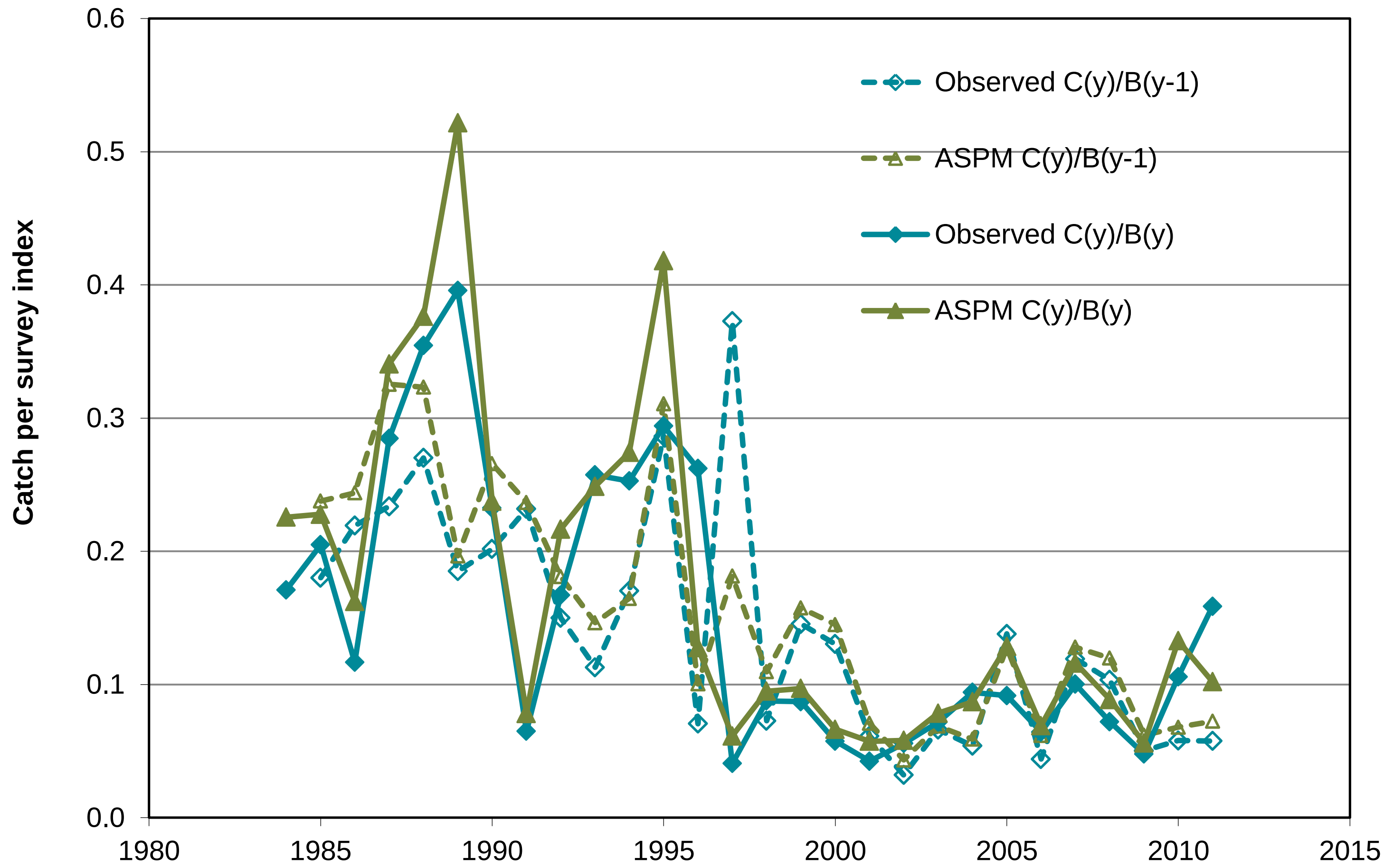
Second of two models: SEVAR

- Stands for “survey/exploitation vector autoregressive”
- Linear time series with p lags in two variables:
  - Survey index (preferably biomass, relative is OK)
  - Ratio of catch to survey index (“exploitation rate”)
- Survey isocline is linear in exploitation rate
  - MSY stock size occurs at ½ equilibrium unexploited
  - Catch = survey × (catch/survey)
- Uses survey index and catch data only
- Used to generate and fit three sets of pseudo-data

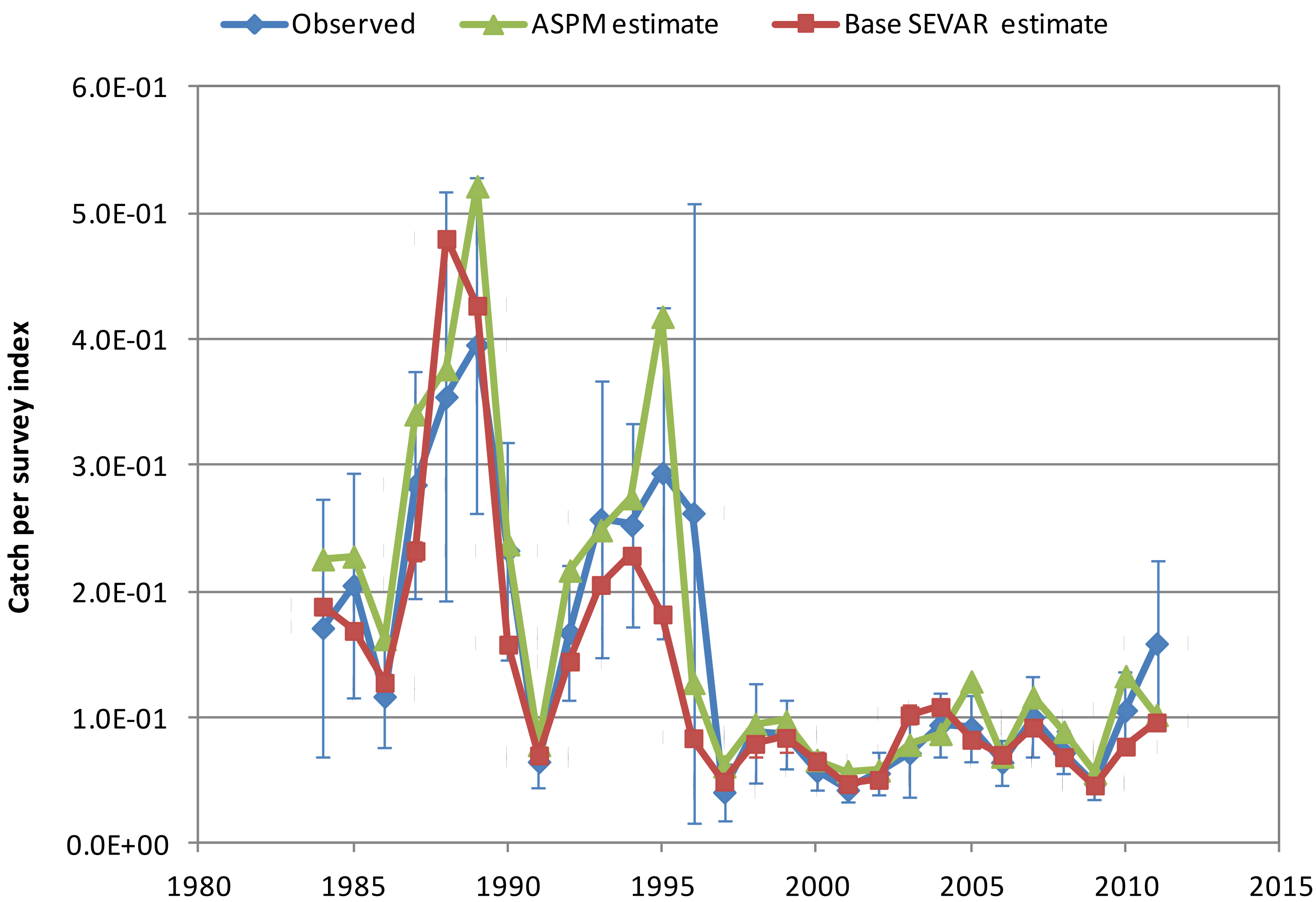
Model fits to observed survey index, B(y)



ASPM fit to observed exploitation rate, C/B



Model fits to observed exploitation: C(y)/B(y)



2×2 factorial bootstrap designs

- Names of designs

	Parametric	Nonparametric
Unconditional	"unc/par"	n/a
Conditional	"con/par"	"con/non"

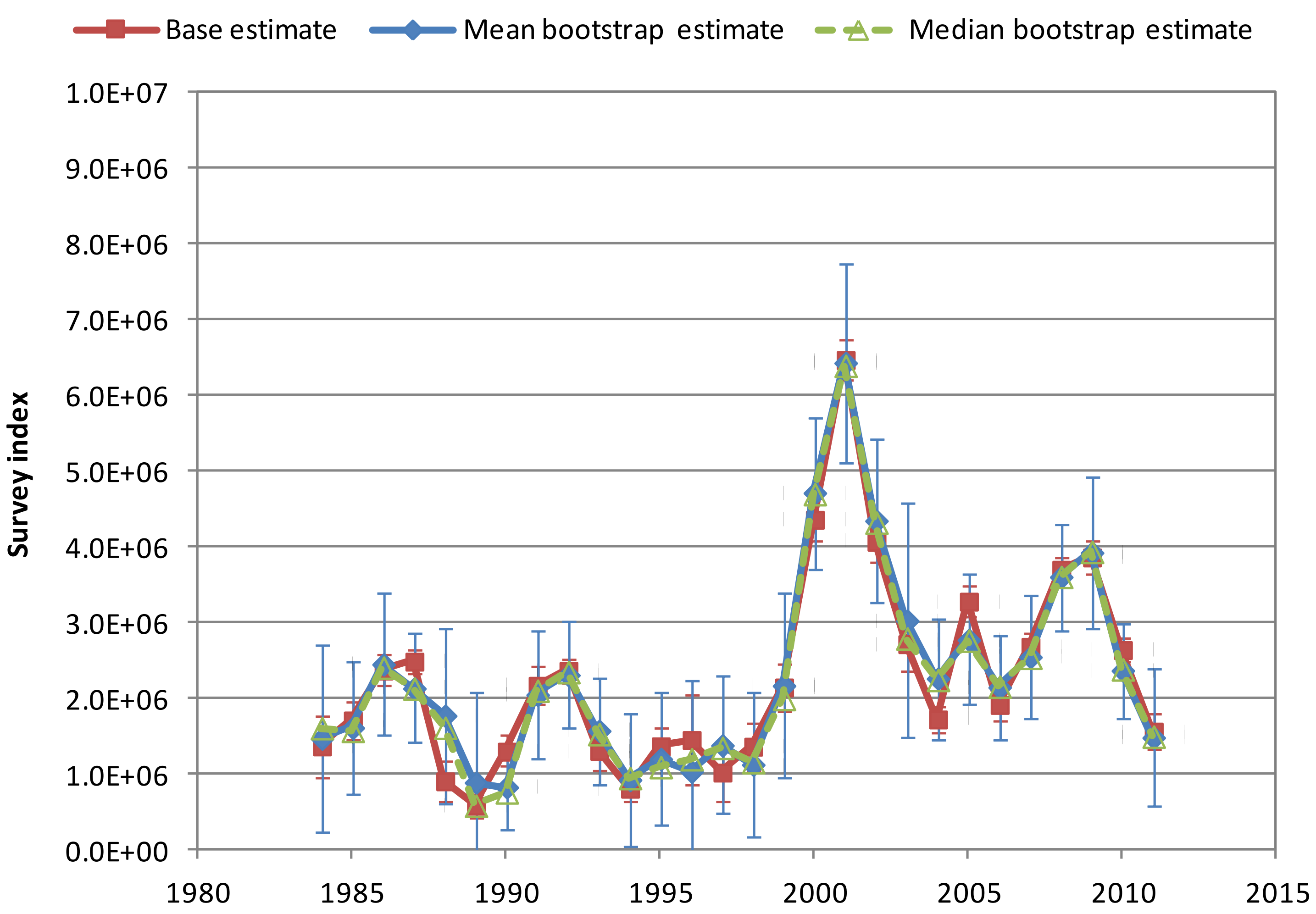
- Error types included

	Parametric	Nonparametric
Unconditional	obs. only	n/a
Conditional	obs. & proc.	obs. & proc.

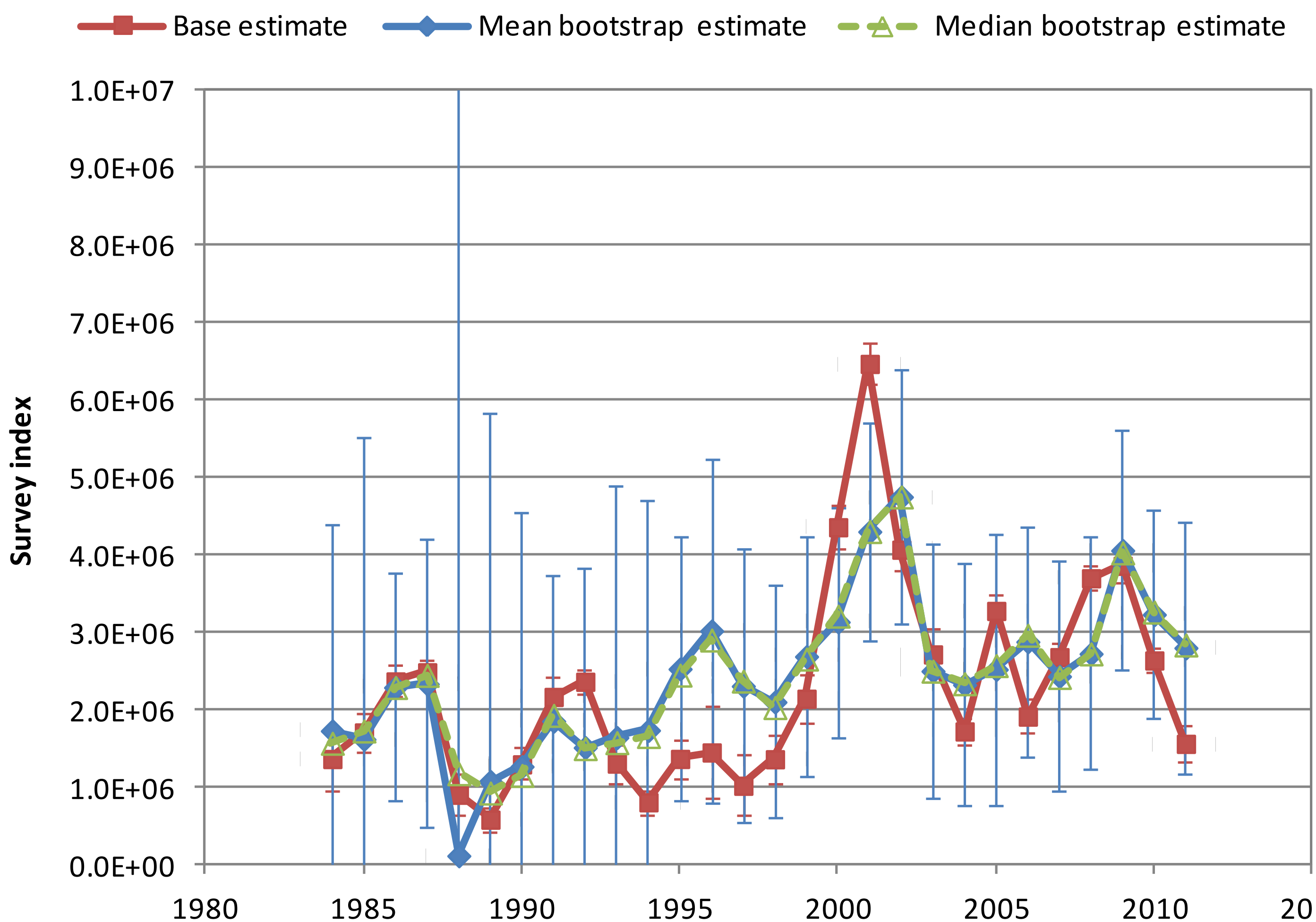
2×2 factorial bootstrap designs

- Unc/par: Random deviates are drawn from the (assumed normal) sampling distributions of total catch and the survey index
- Con/par: Random deviates are drawn from the bivariate normal distribution used to define the likelihood in the SEVAR model
- Con/non: Bivariate residuals from the SEVAR model are standardized, sampled with replacement, "de-standardized" as appropriate for each year in the time series, then added to the model's predicted value for that year in the time series

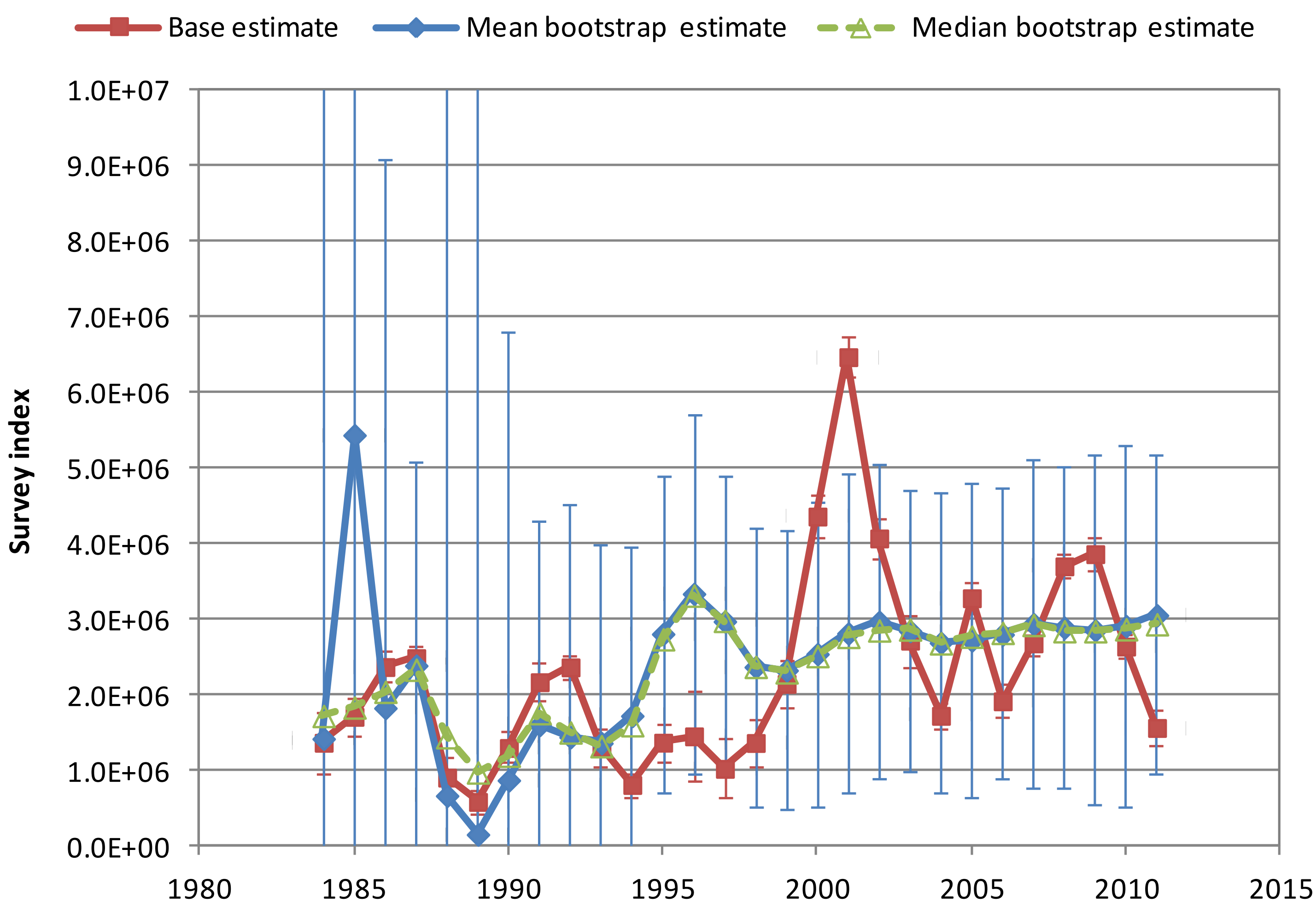
Unc/par bootstrap vs. base (survey index)



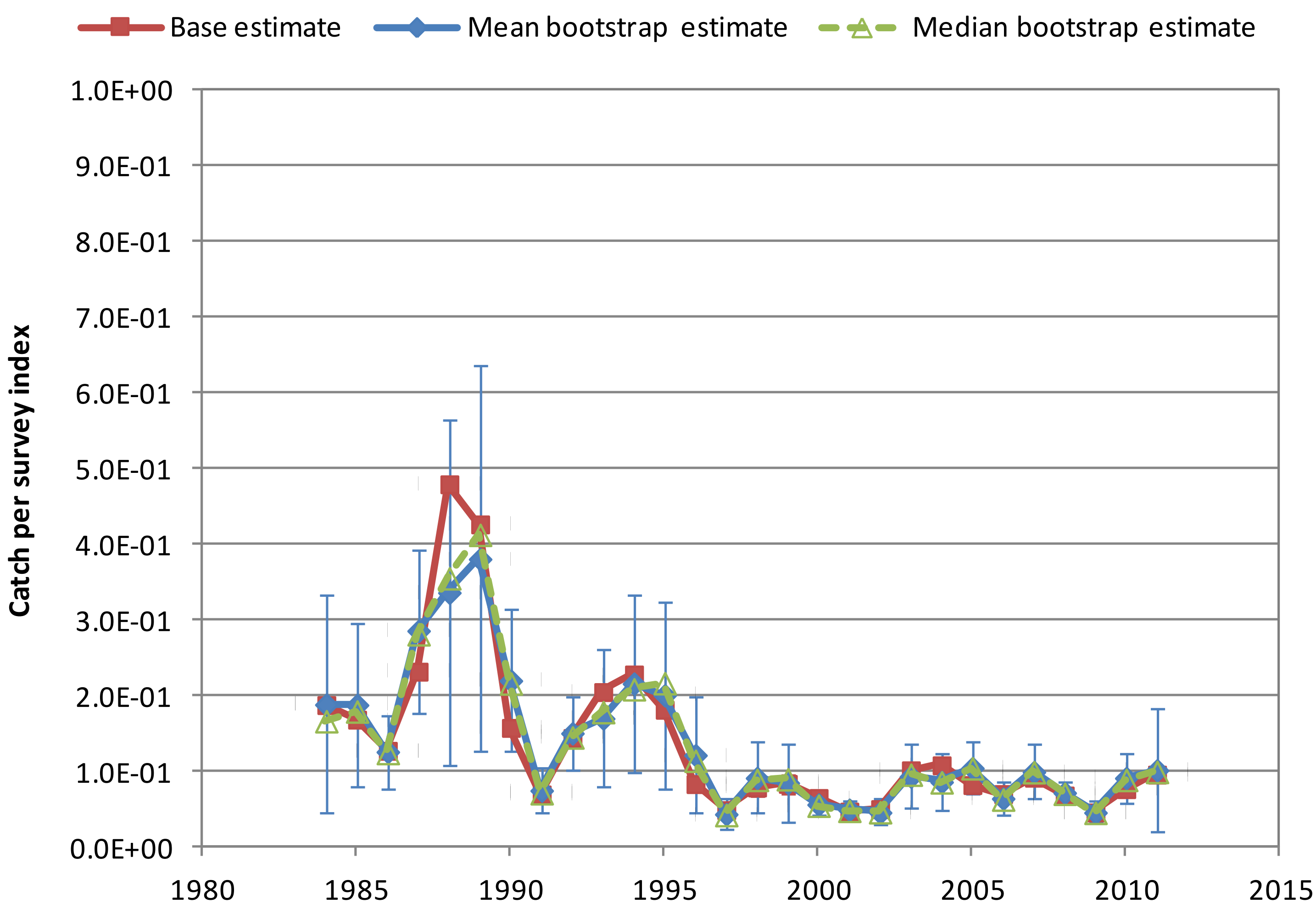
Con/par bootstrap vs. base (survey index)



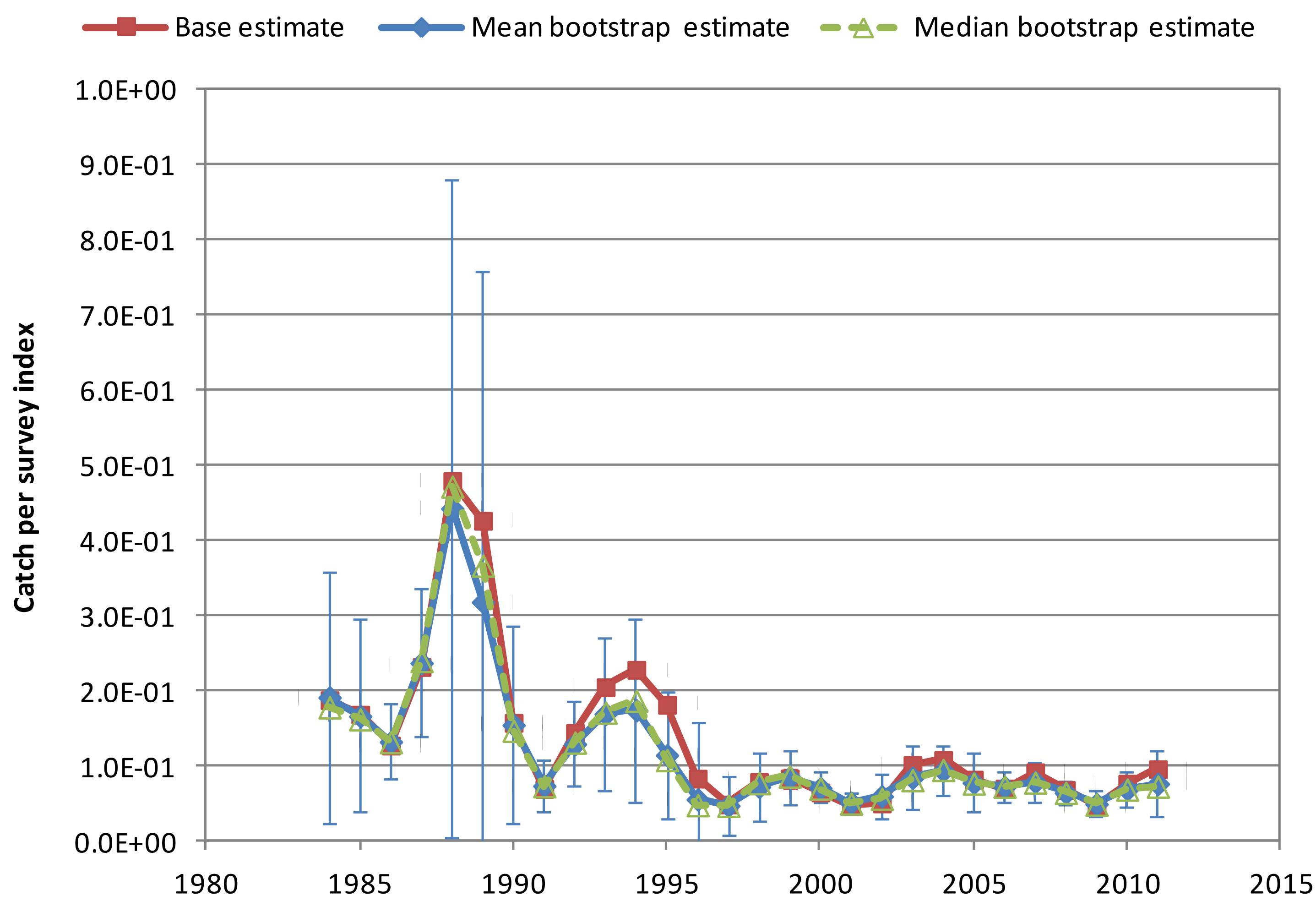
Con/non bootstrap vs. base (survey index)



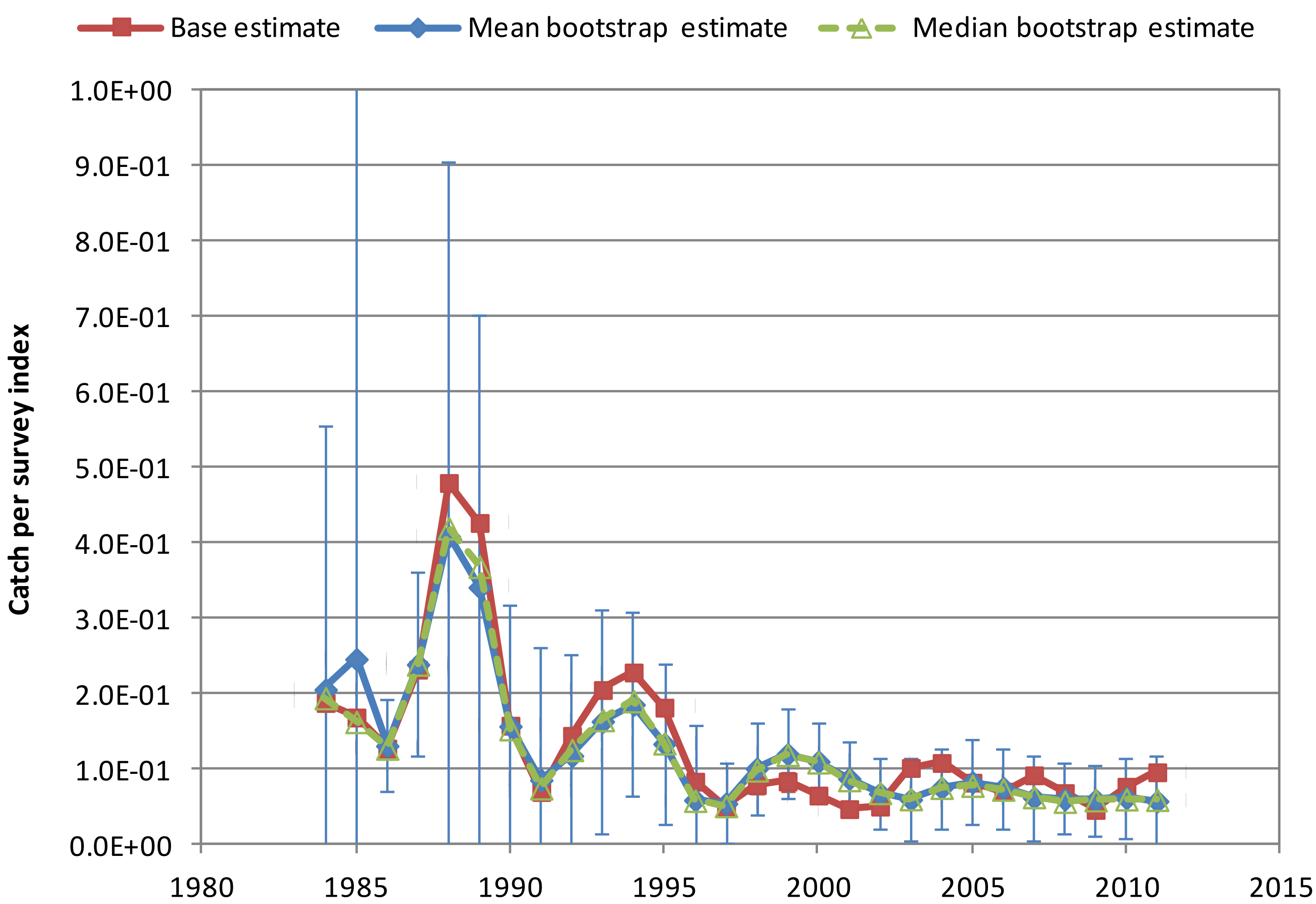
Unc/par bootstrap vs. base (exploitation)



Con/par bootstrap vs. base (exploitation)



Con/non bootstrap vs. base (exploitation)

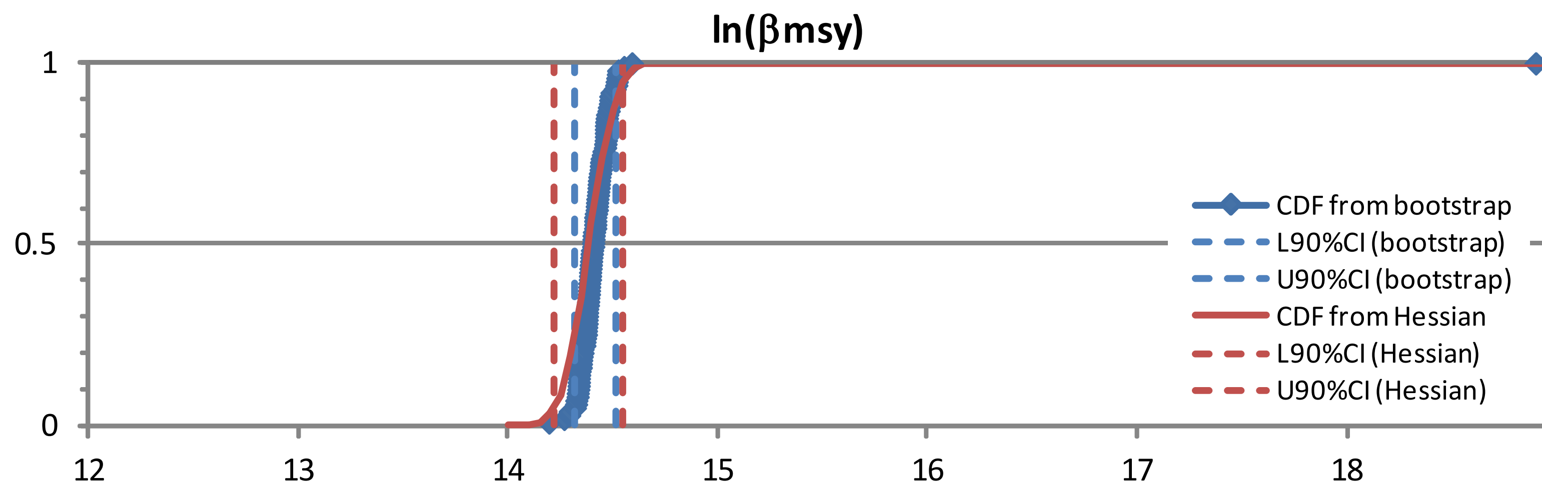


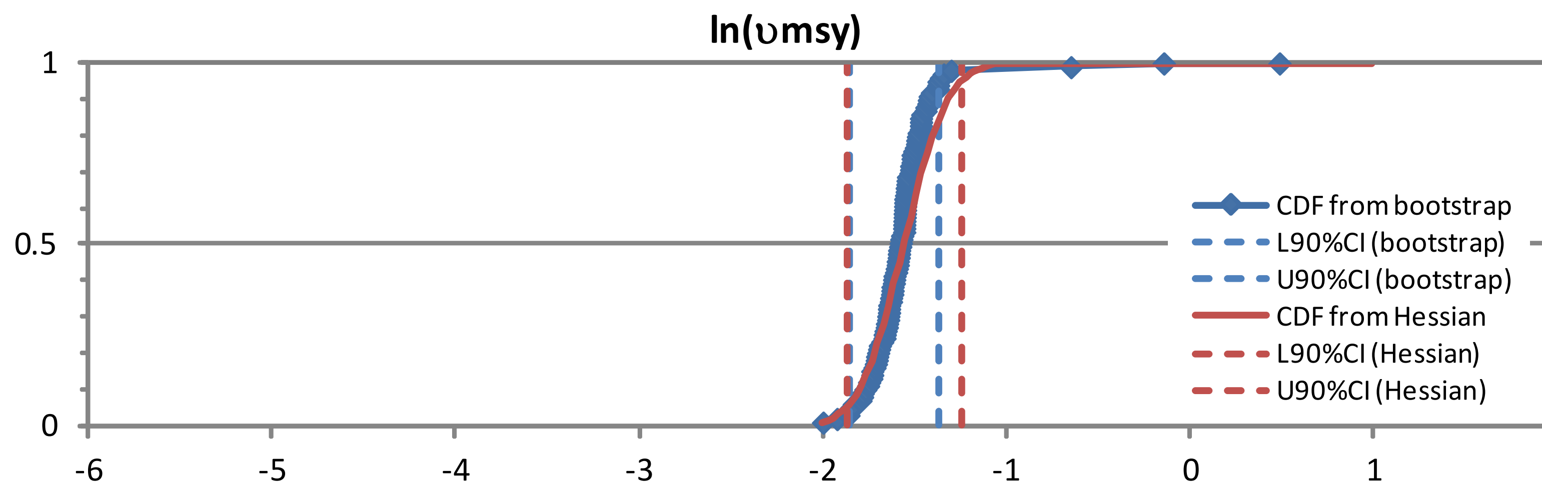
CDFs of key management quantities

- Three slides (one for each bootstrap type) for three pairs of management quantities:
  - Survey biomass and exploitation rate at MSY
  - MSY and projected (1-yr-ahead) survey biomass
  - "Depletion" (projected survey biomass relative to equilibrium unfished survey biomass) and the "overfishing level" (1-yr-ahead catch at MSY rate)
- All management quantities are on natural log scale
- For each management quantity, horizontal axis range is the same for each bootstrap type

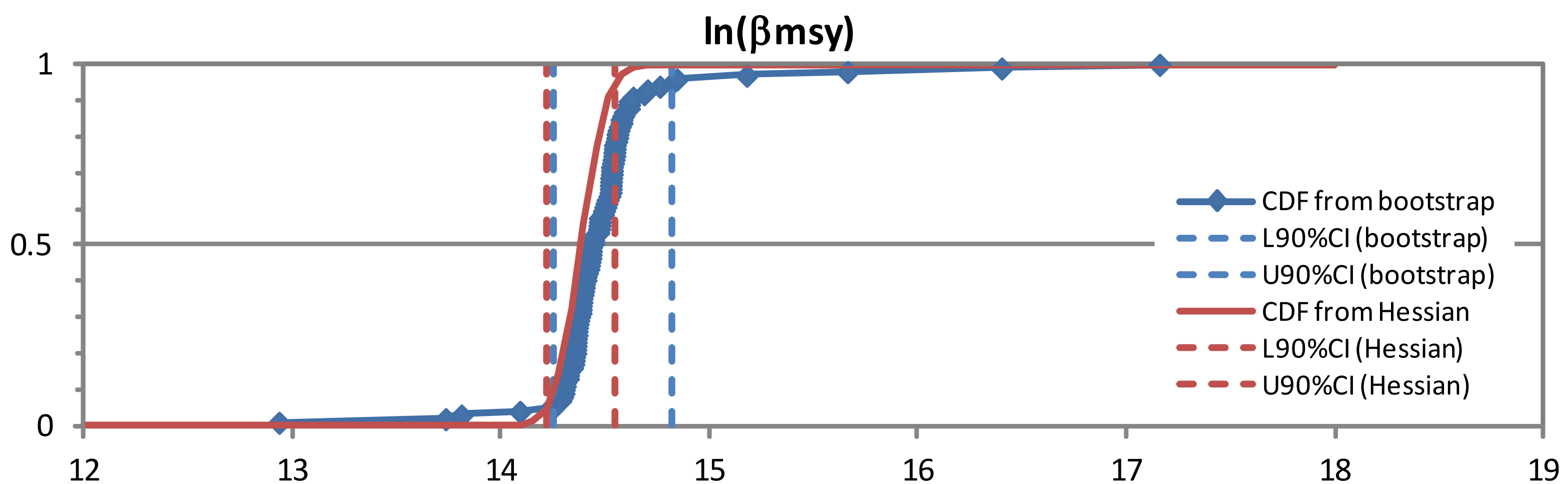
NOAA FISHERIES

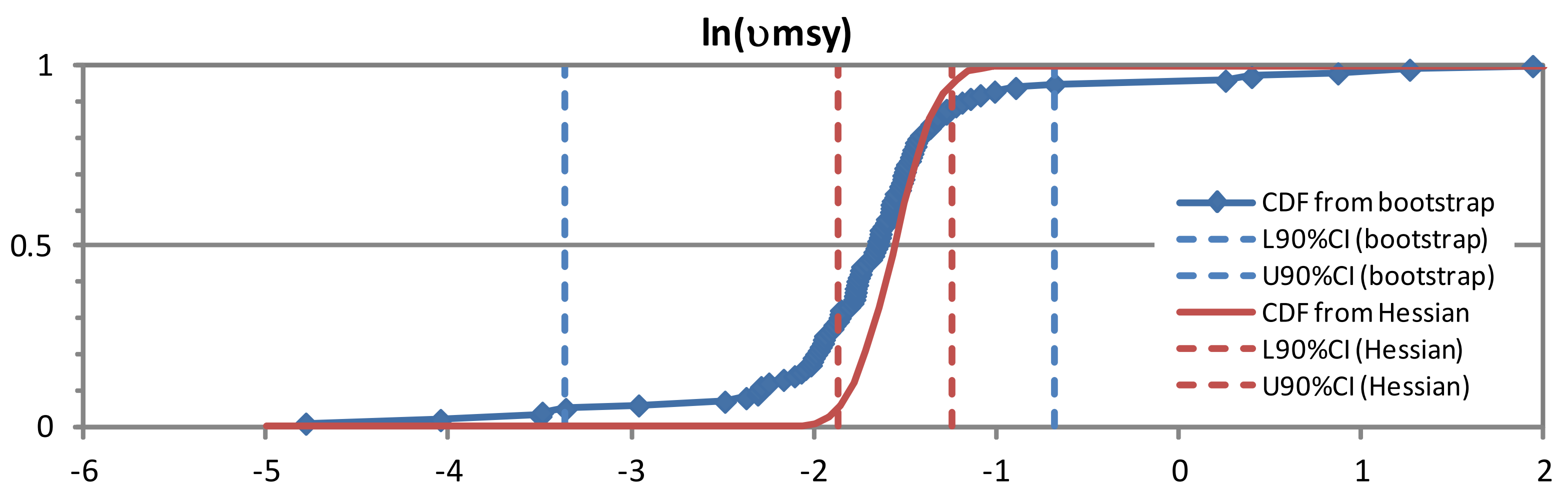
Biomass and exploitation at MSY: unc/par



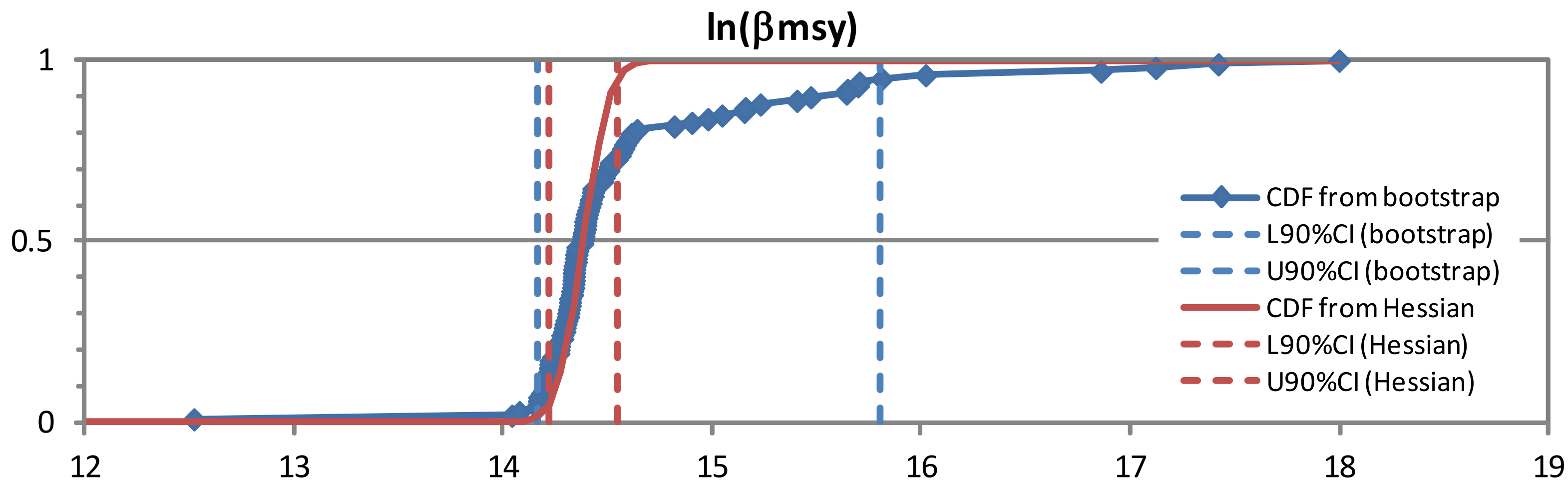


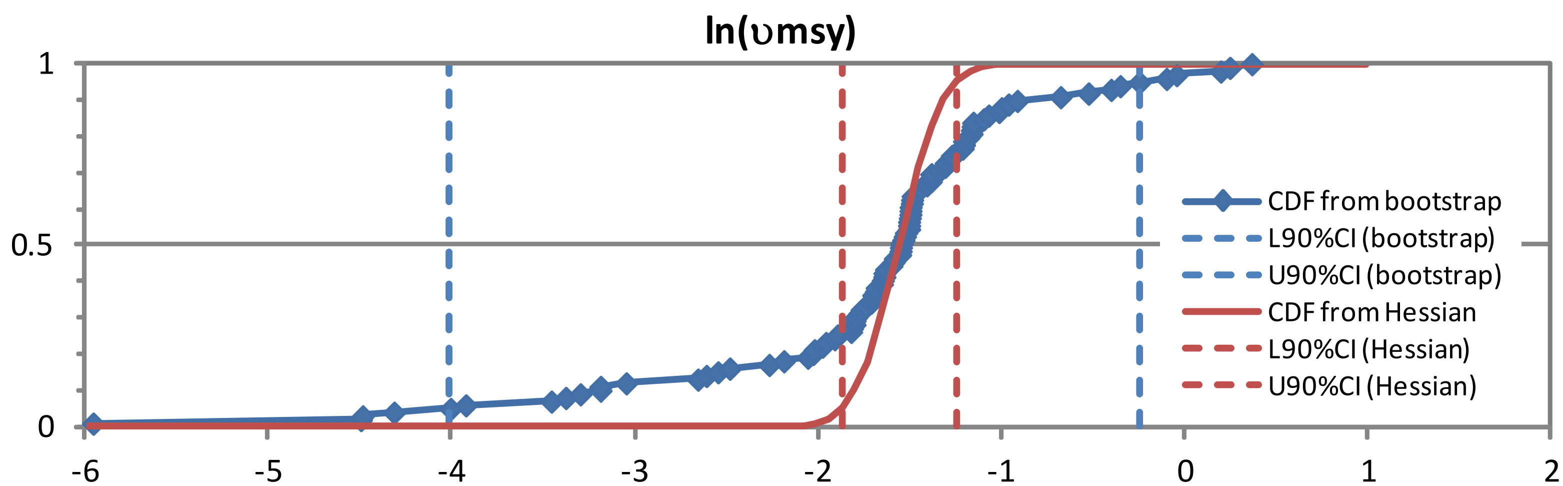
Biomass and exploitation at MSY: con/par



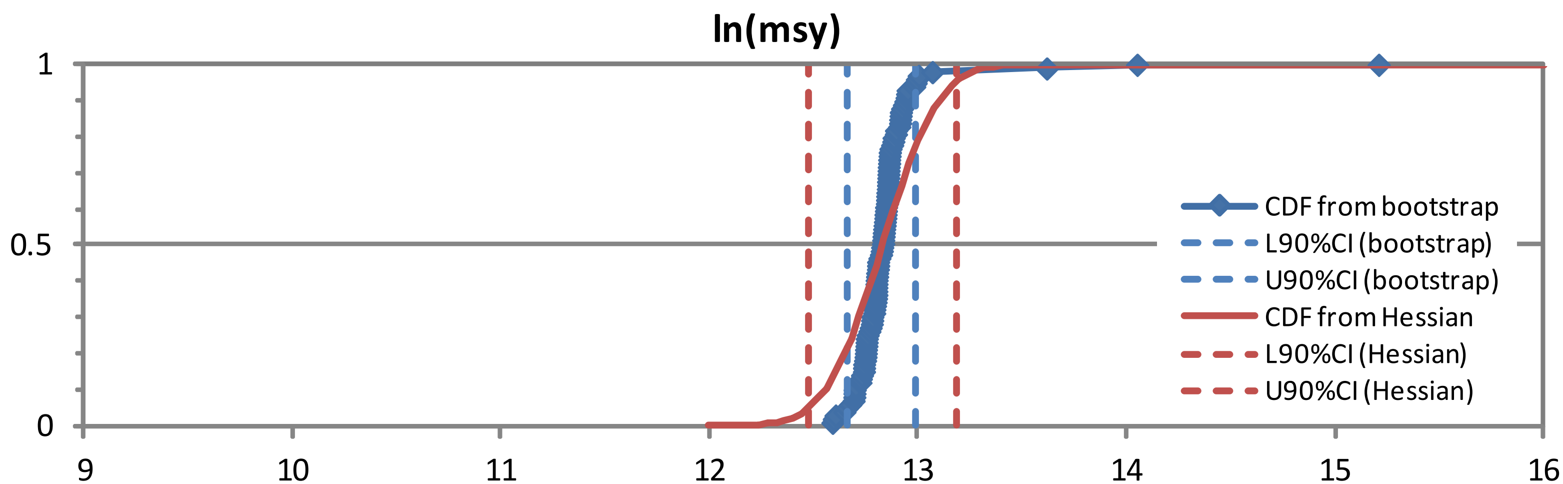


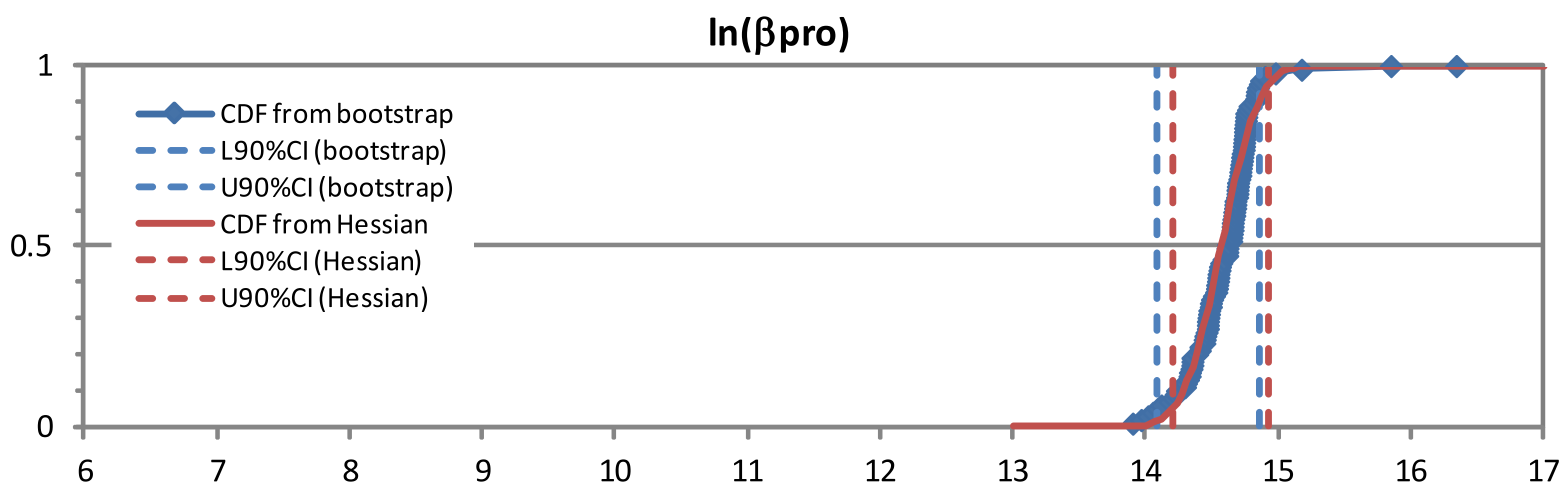
Biomass and exploitation at MSY: con/non



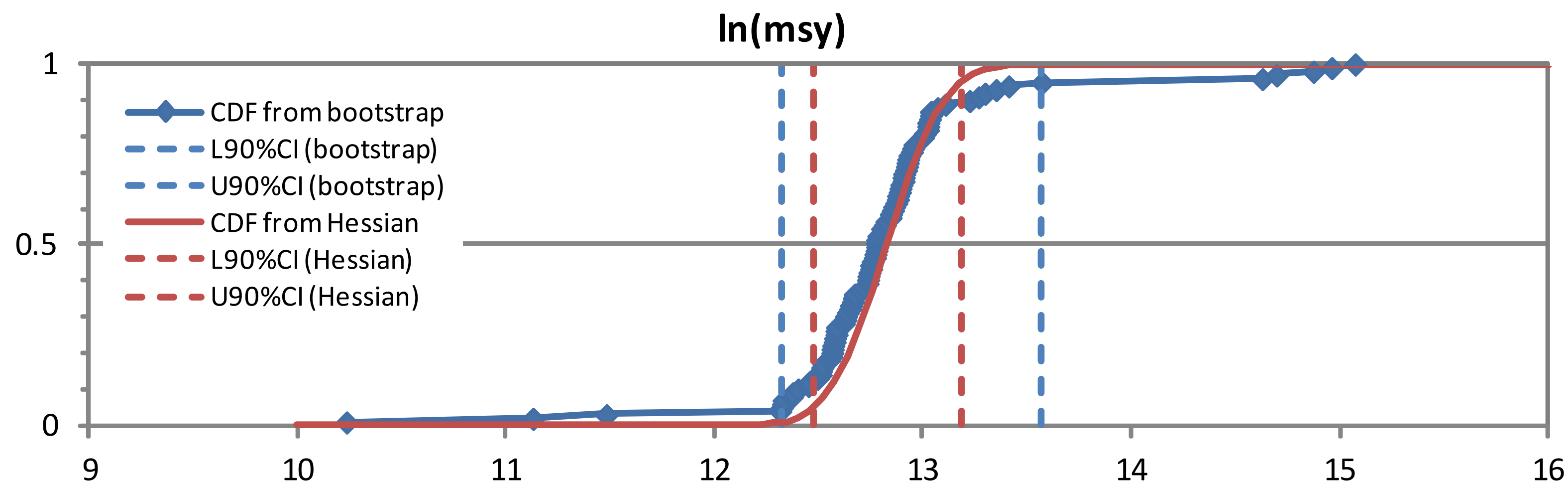


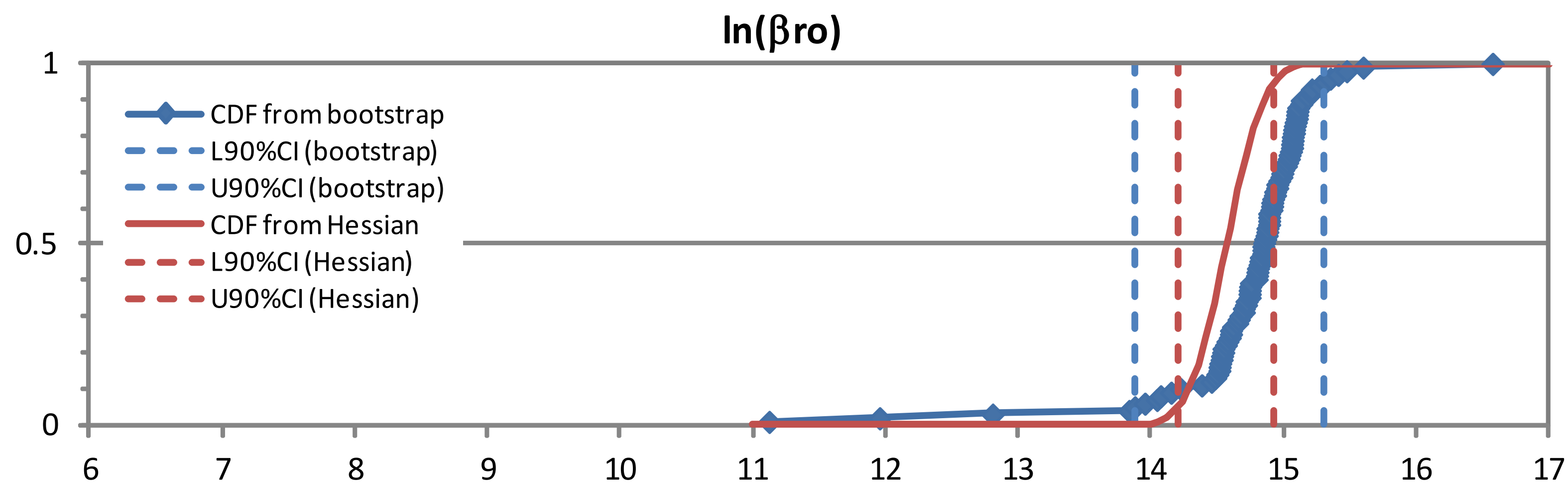
MSY and projected biomass: unc/par



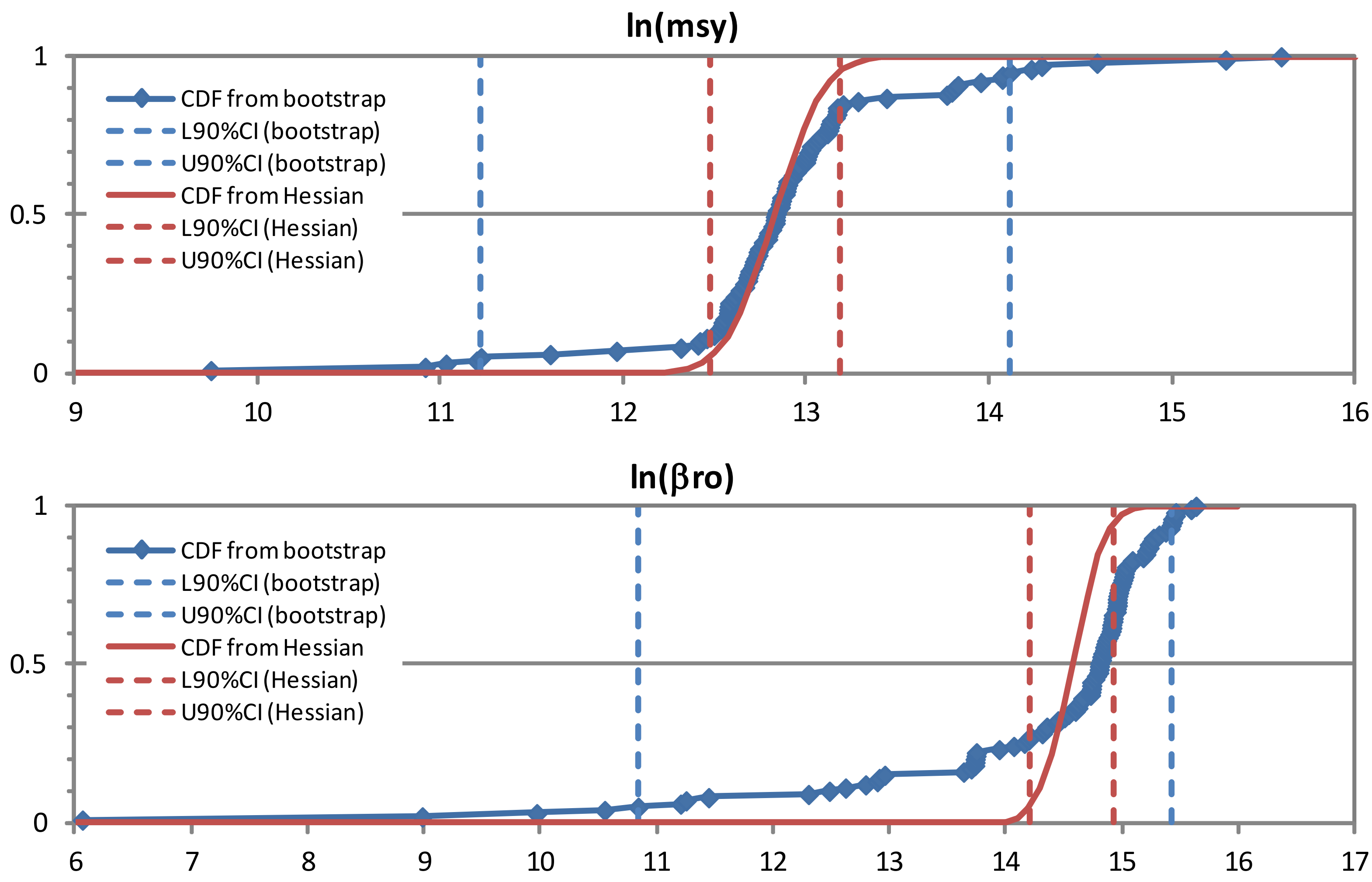


MSY and projected biomass: con/par

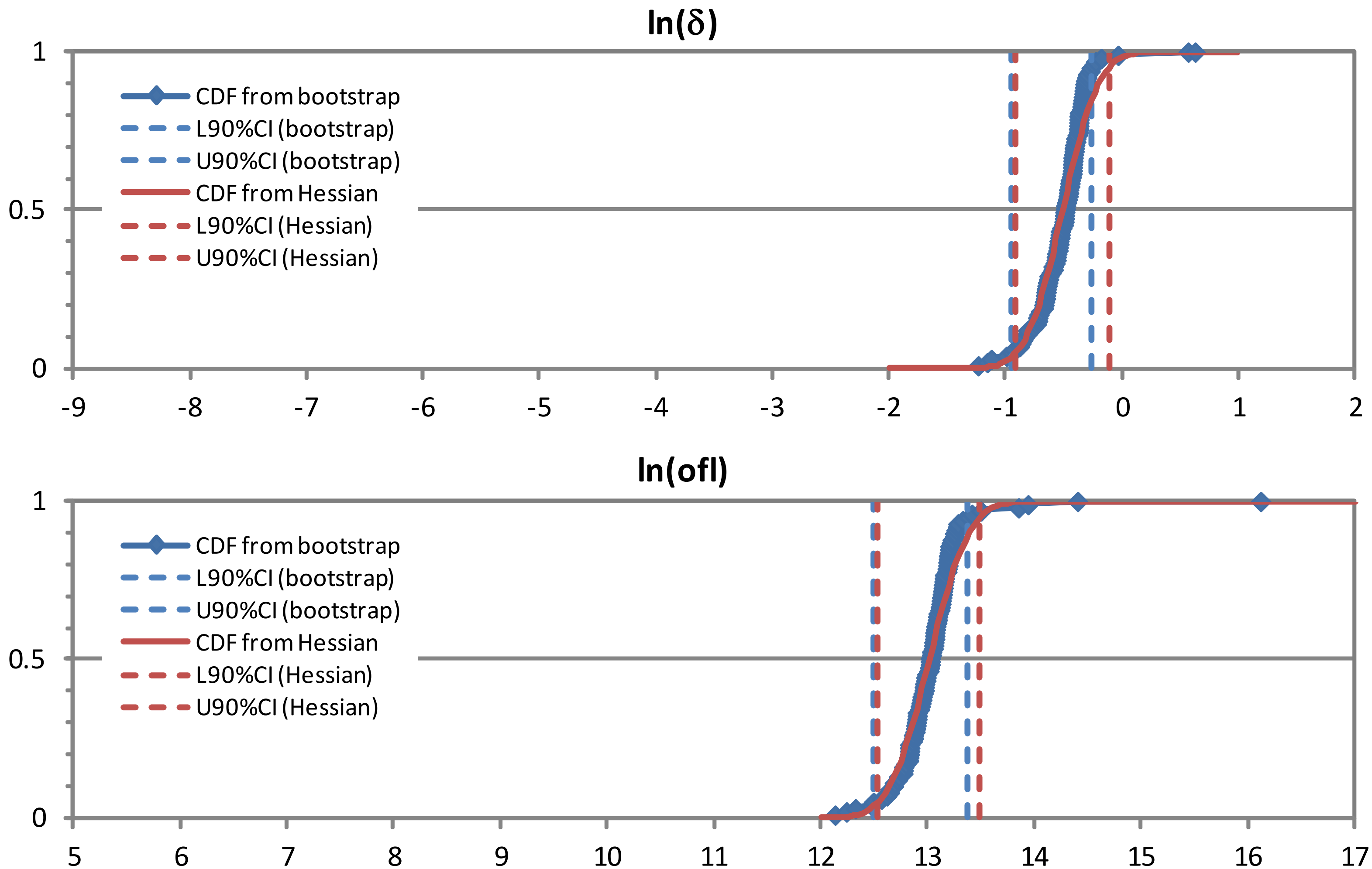




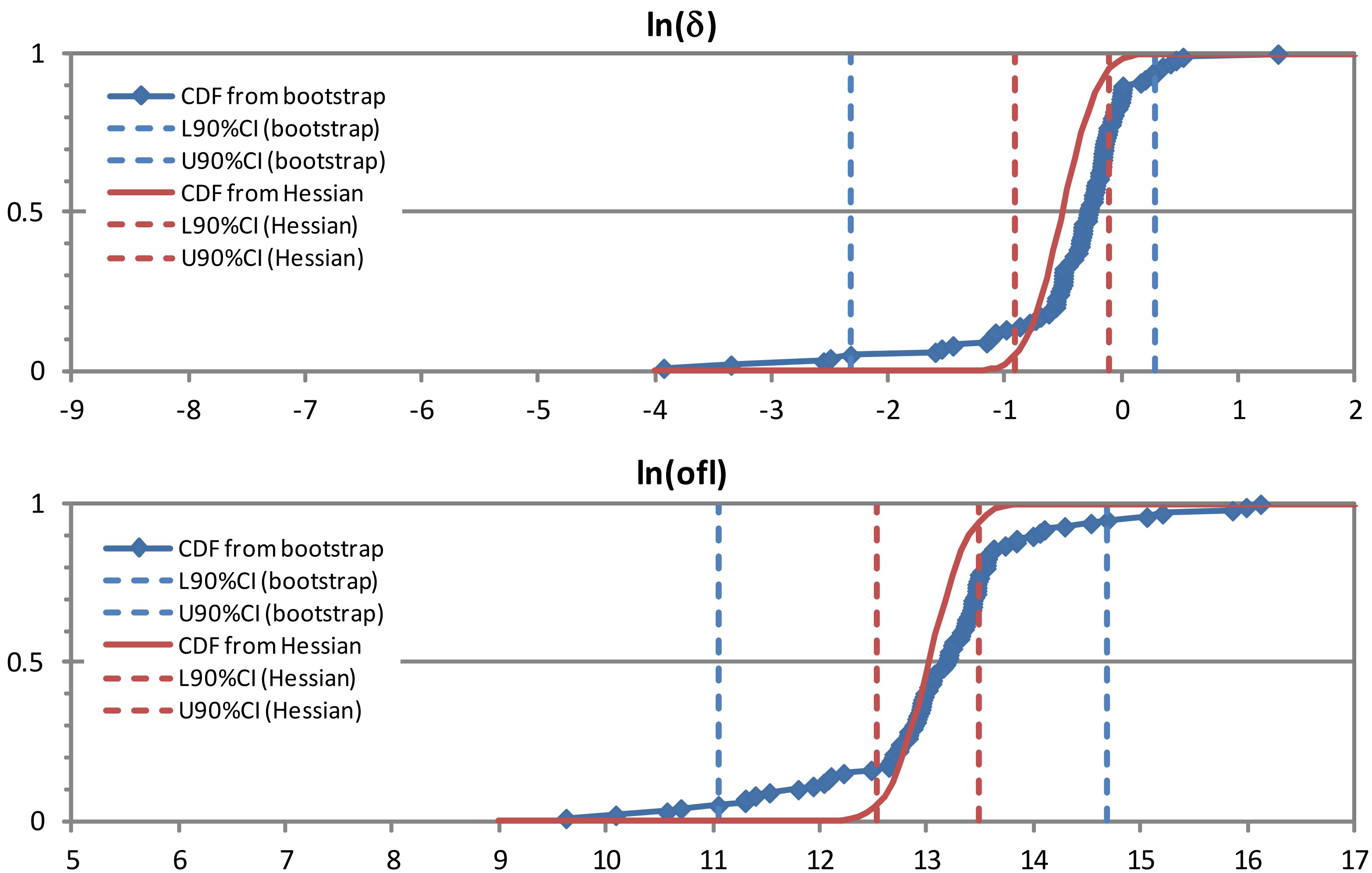
MSY and projected biomass: con/non



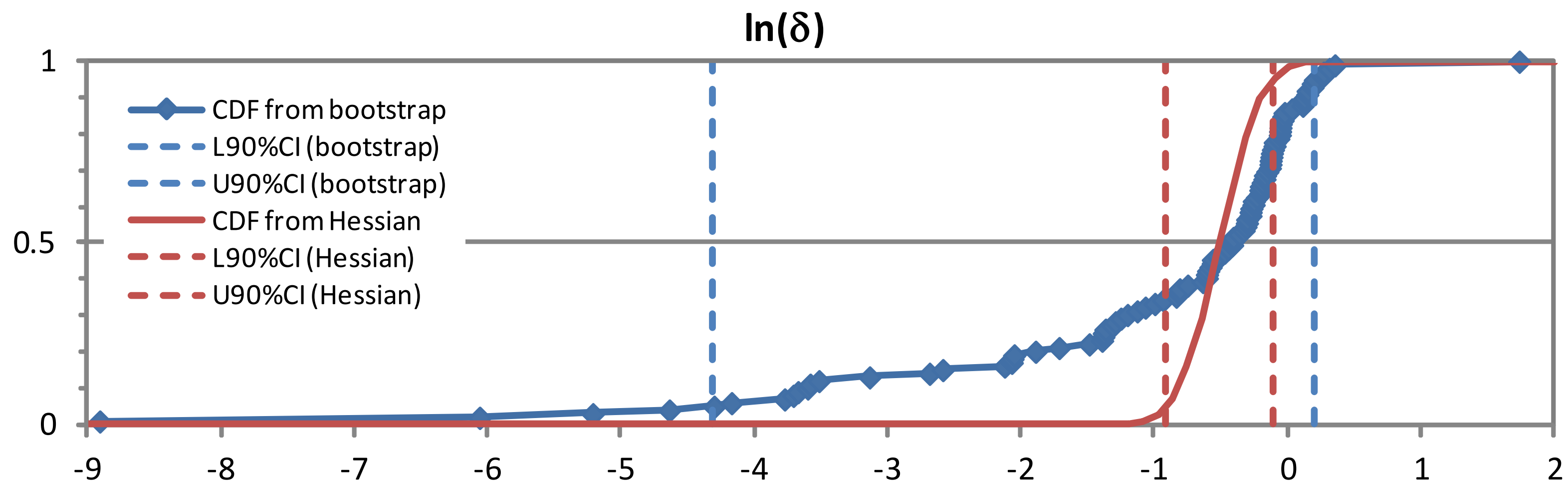
Depletion and overfishing level: unc/par

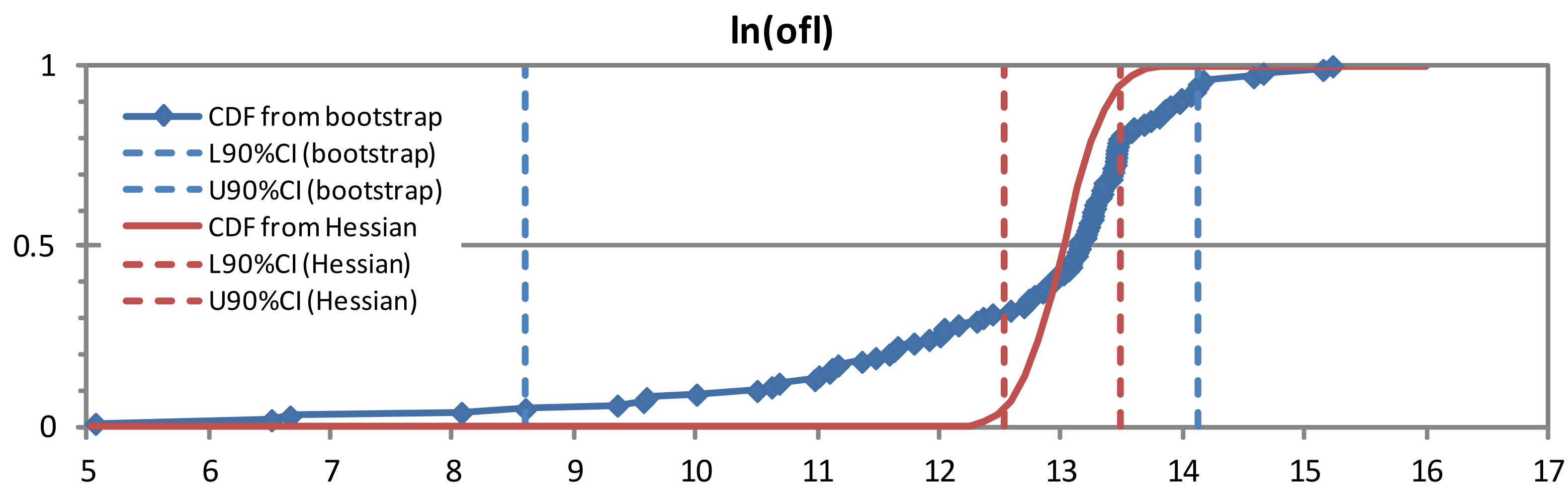


Depletion and overfishing level: con/par



Depletion and overfishing level: con/non





Potential discussion topics

- Can time series models provide meaningful management advice?
- Should use of time series models be restricted to data-moderate situations only, or can they provide a useful counterpart to full, age-structured models?
- Are distributions based on bootstraps preferable to Hessian approximations?
- What are the implications of performance differences between unc/par, con/par, and con/non?

Acknowledgements

- Keith Criddle
- George Sugihara and others at 2012 NLTS workshop
- Jim Ianelli
- Steve Barbeaux
- Jon Deroba
- Doug Butterworth, Steve Cadrin, Rick Methot

Lessons Learned from a Stock Assessment Simulation Study

Athol Whitten, Carey McGilliard, Juan Valero
University of Washington and **NOAA Fisheries**

For the World Conference on Stock Assessment Methods
Boston, 2013

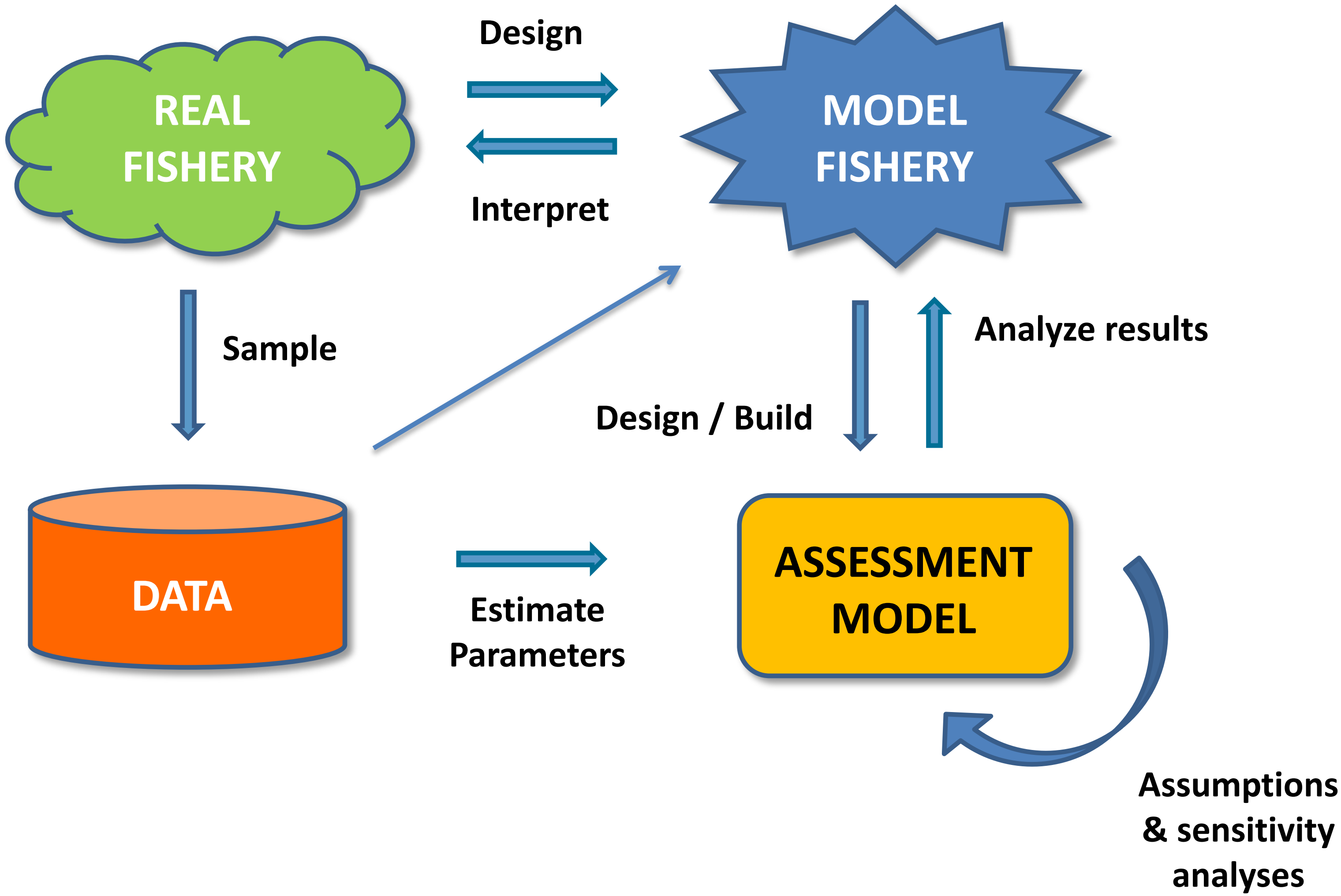
Acknowledgements

Co-authors: Sean Anderson, Curry Cunningham, Felipe Hurtado Ferro, Kelli Johnson, Roberto Licandeo, Cole Monnahan, Melissa Muradian, Kotaro Ono, Cody Szuwalksi, and Katyana Vertpre

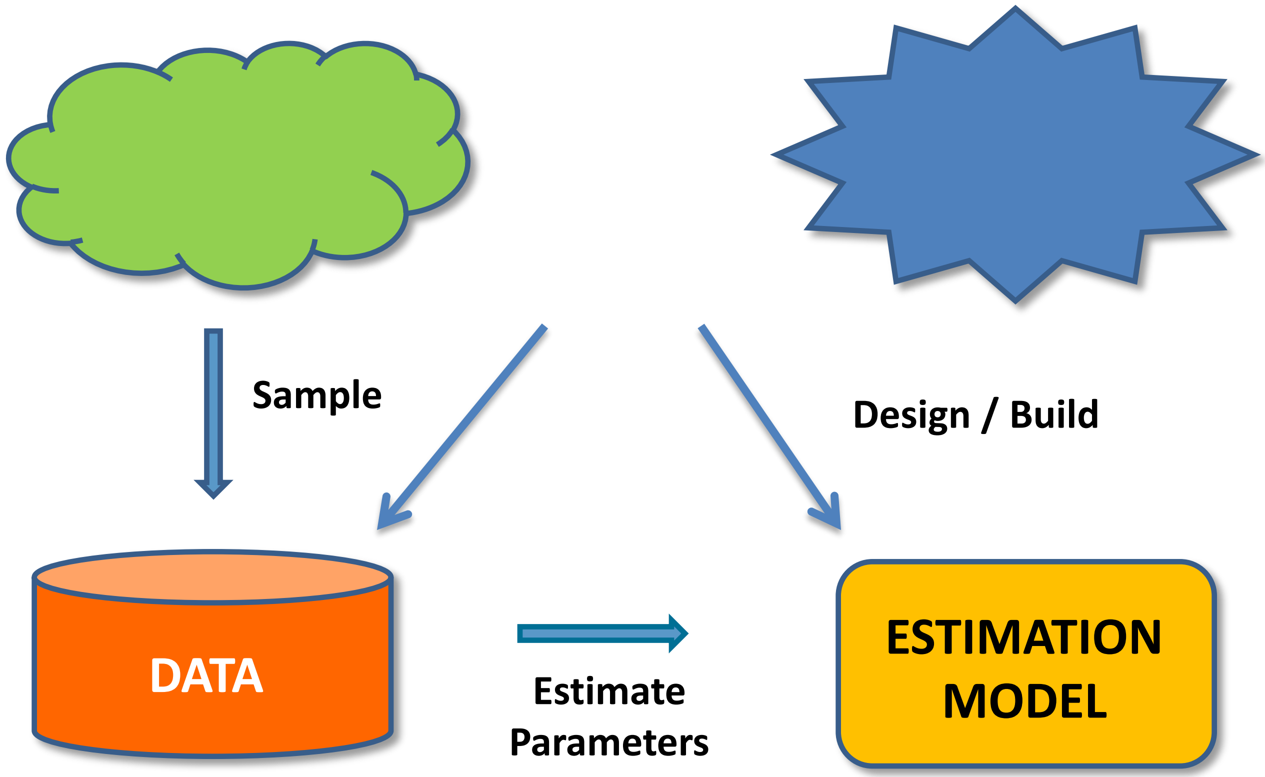
Advised by: André Punt, Richard Methot, and James Ianelli

People from: University of Washington, NOAA Fisheries, University of British Columbia, Simon Fraser University, and CAPAM (Centre for the Advancement of Population Assessment Methodology)

And help from: Ian Taylor and Jim Thorson



REAL FISHERY
Design
Interpret
MODEL FISHERY
Sample
Analyze results
Design / Build
DATA
Estimate Parameters
ASSESSMENT MODEL
Assumptions & sensitivity analyses


Sample
Design / Build
DATA
Estimate
Parameters
ESTIMATION
MODEL

SPECIES	Cod	Flatfish	Sardine
MODELS	Conditioning	Operating	Estimating
QUESTIONS	Natural Mort	Data Quality	Retrospective
AIMS	Relevance	Realism	Reproducibility

SPECIES	Cod	Flatfish	Sardine
MODELS	Conditioning	Operating	Estimating
QUESTIONS	Natural Mort	Data Quality	Retrospective
AIMS	Relevance	Realism	Reproducibility

SPECIES

Cod		Long lived, slow growing, low M
Flatfish		Medium life span, medium growth, low M
Sardine		Short lived, fast growing, high M

MODELS

All achieved in a single modelling framework: *Stock Synthesis*

Conditioning	Based on existing stock assessment models
Operating	Simplified version of conditioning models
Estimating	'Standard' stock assessment approaches

QUESTIONS

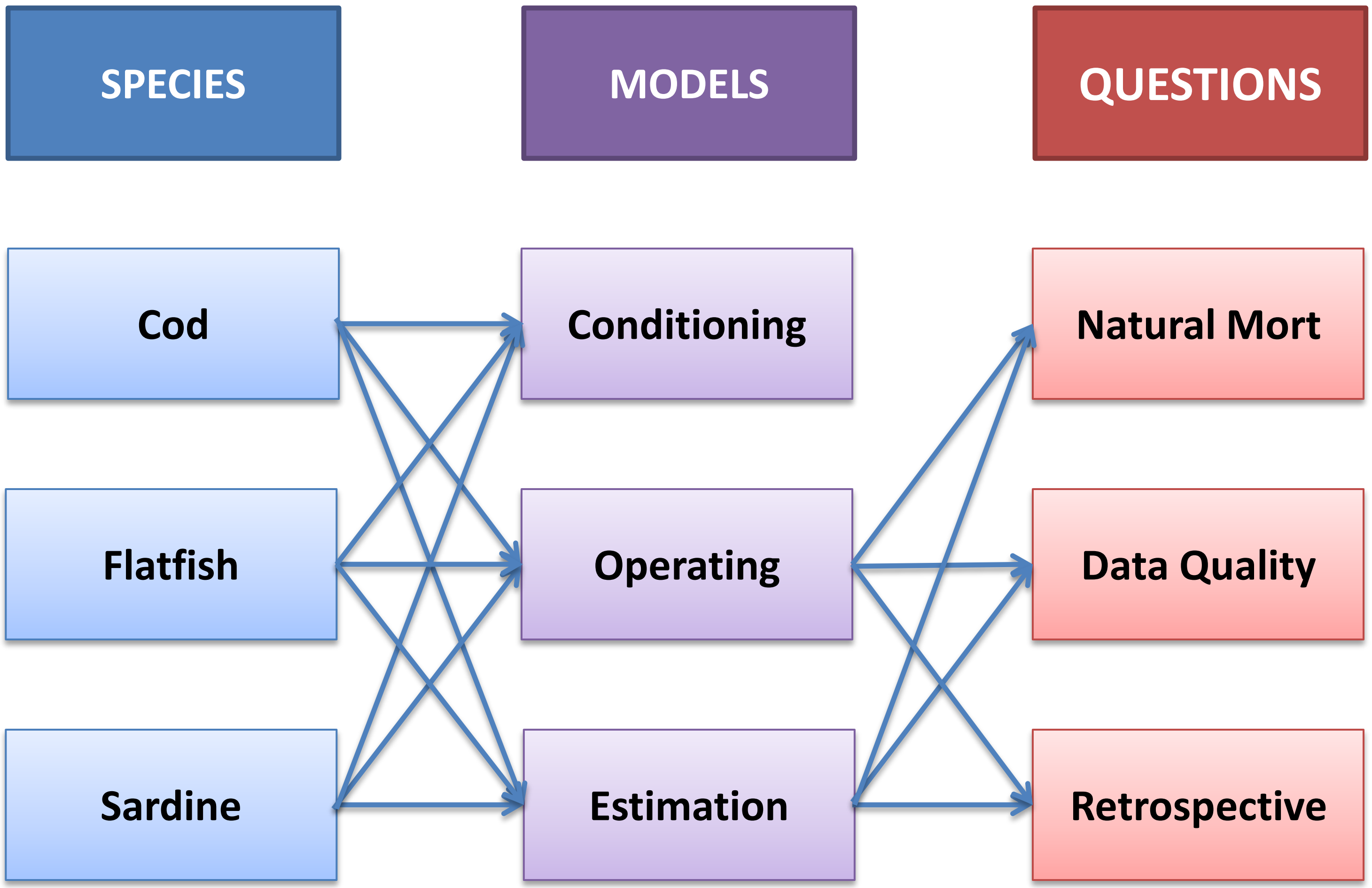
Natural Mort	Dealing with Time-Varying M	Kelli Johnson
Data Quality	Which types of data are essential?	Kotaro Ono
Retrospective	What produces retrospective patterns?	Felipe Hurtado Ferro

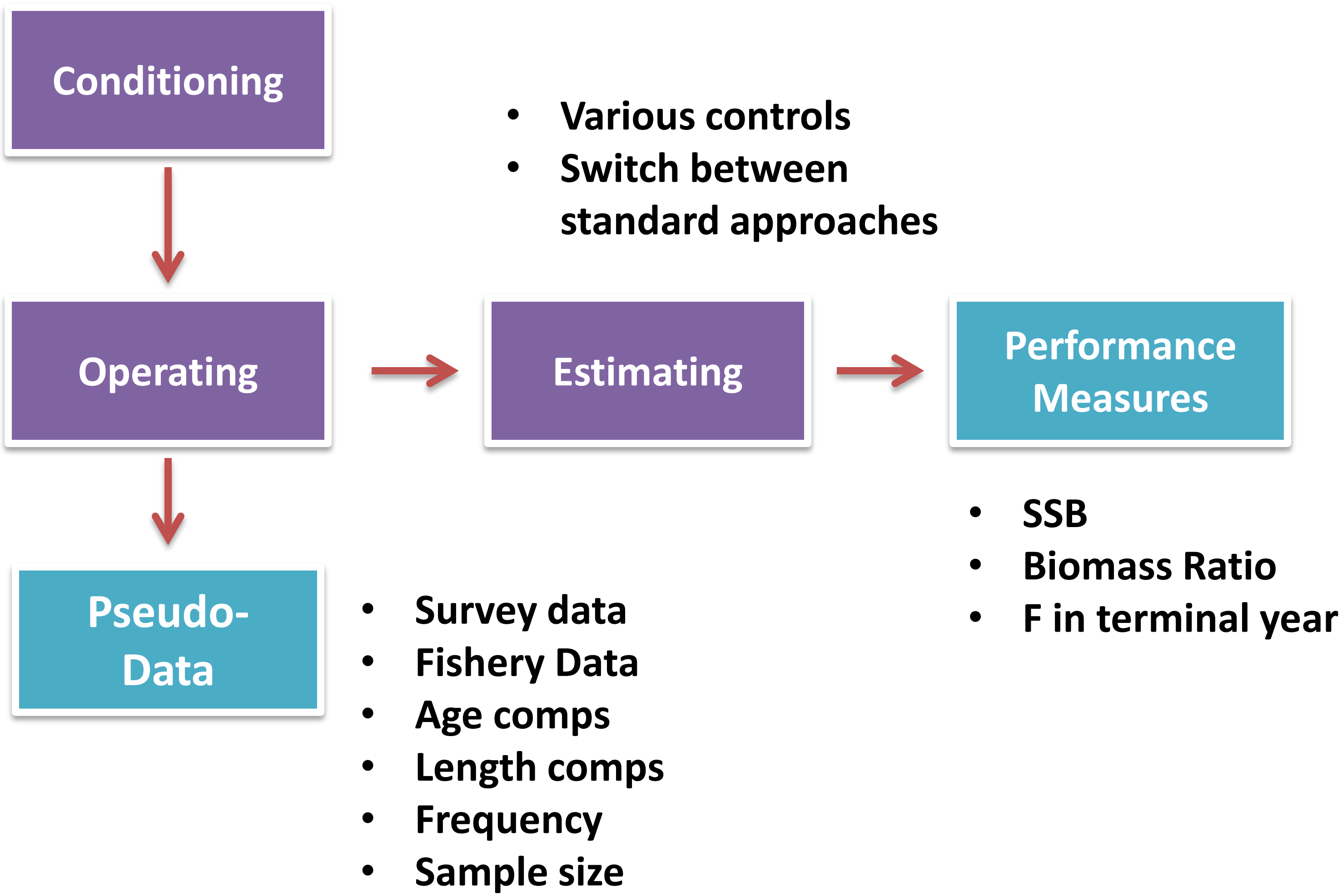
QUESTIONS

Natural Mort

Data Quality

Retrospective


SPECIES
MODELS
QUESTIONS
Cod
Conditioning
Natural Mort
Flatfish
Operating
Data Quality
Sardine
Estimation
Retrospective


Conditioning
Various controls
Switch between standard approaches
Operating
Estimating
Performance Measures
Pseudo-Data
Survey data
Fishery Data
Age comps
Length comps
Frequency
Sample size
SSB
Biomass Ratio
F in terminal year

RESULTS

Time-varying M	Majority rules: Keep M fixed at long term value
Data quality and quantity	Survey composition data are important!
Retrospective Patterns	These can be generated with time-varying parameters of many kinds

Lessons Learned

A good study design is essential…

Test code, test assumptions, self-test models.

Standardized and versioned code enables great progress.

With collaborative research we can achieve great things!

Thanks!

The a4a Initiative
Simulation testing

Ernesto Jardim*
Marco Ferretti
Colin Millar

JRC - Unit of Maritime Affairs, Fishreg
*** ernesto.jardim@jrc.ec.europa.eu**

European Commission

assessment for all (a4a)

Long term objective - To have a group of standard methods that can be applied rapidly to a large number of stocks, without requiring a strong statistical technical background, but making use of the technical knowledge on the fisheries, stocks and ecosystem characteristics.

European Commission

Simulation:

- *Test how well the model rebuilds the truth under a range of conditions.*
- *Test “automatic mode”.*

[With R/FLR (methods, data structures, parallel computing, easy data analysis, repeatability)]

European Commission

Approach:

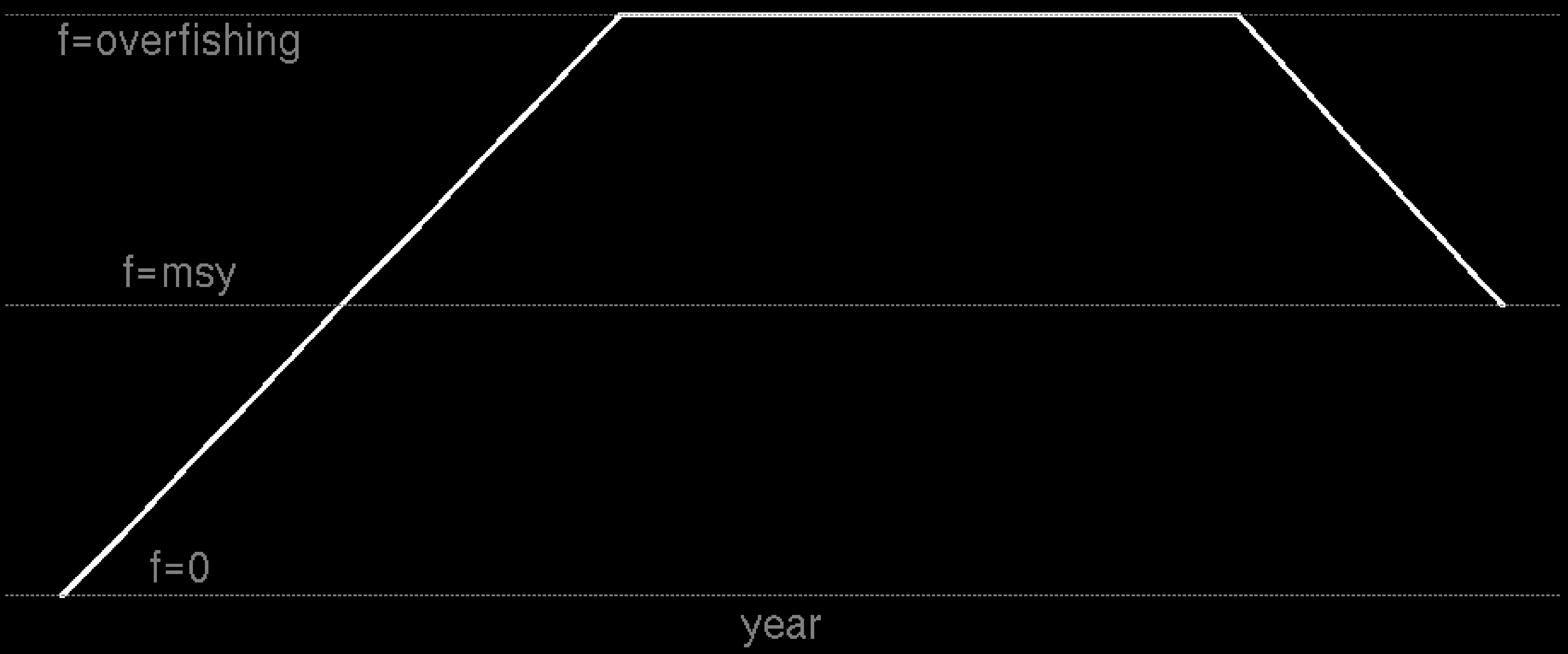
- *Generate OMs based on biology and exploitation characteristics.*
- *Add observation error.*
- *Fit models.*
- *Compare with the simulated data.*

European Commission

Algorithm step 01 – get life history parameters

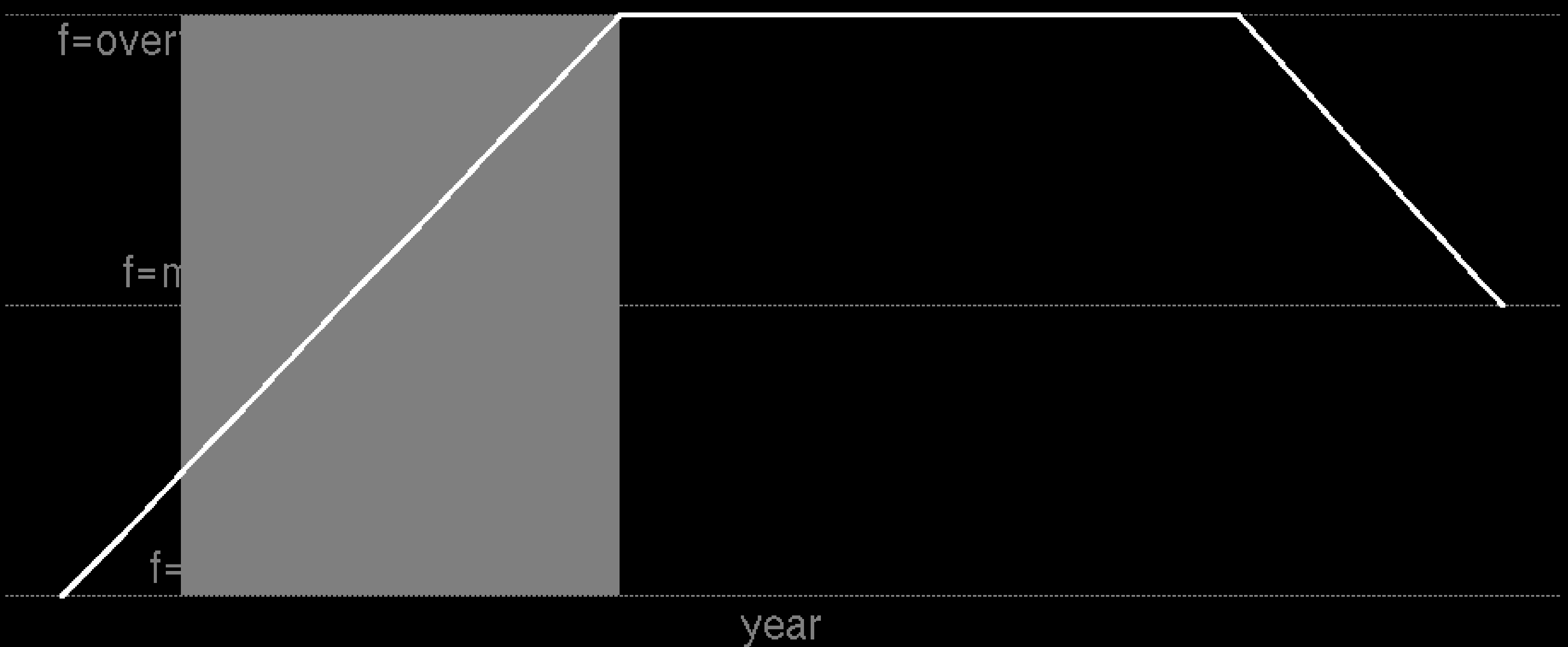
- *webscrap fishbase for life history parameters [**a, b, Linf, K, T0, L50, a50**]*
- *two S/R models beverton & holt or ricker with two steepness values 0.6 or 0.8.*
- *build coherent population dynamics under no-exploitation*
- *1053 species*

Algorithm step 02 - simulate exploitation

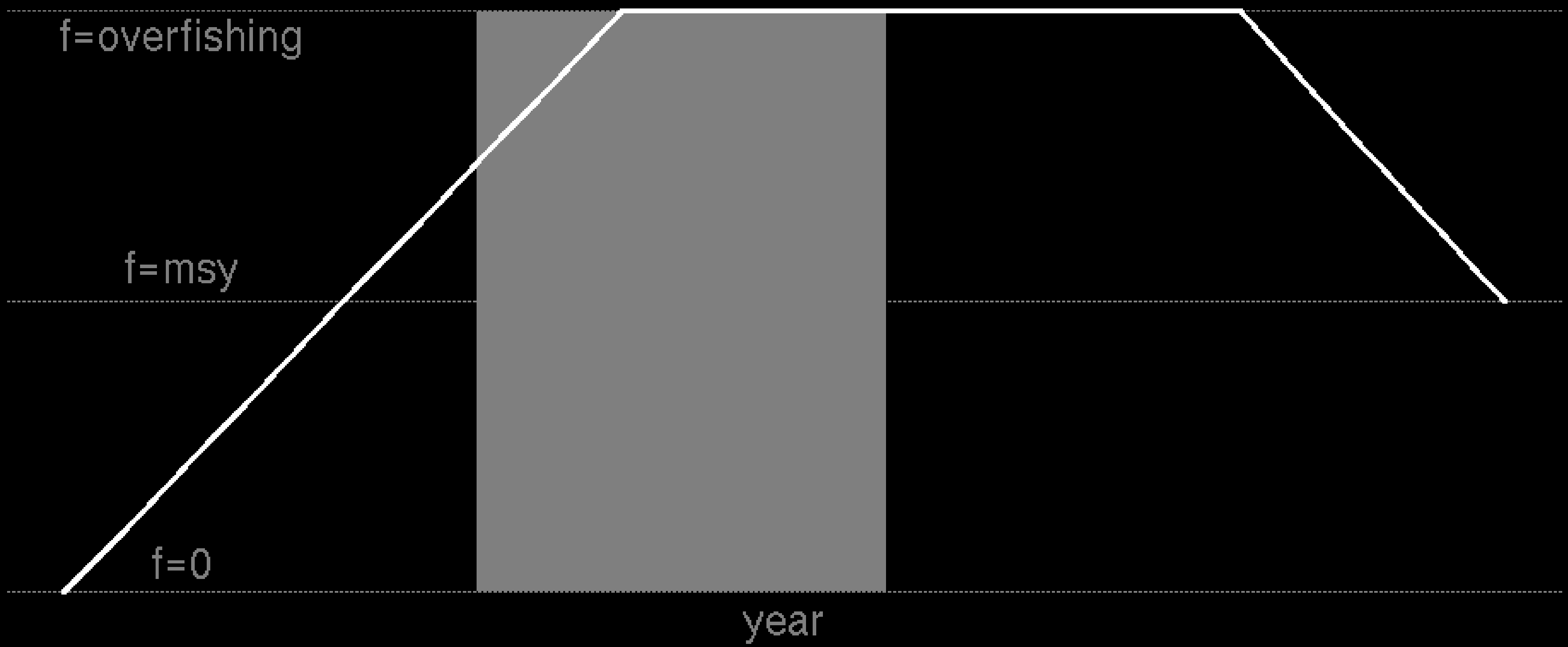


European Commission

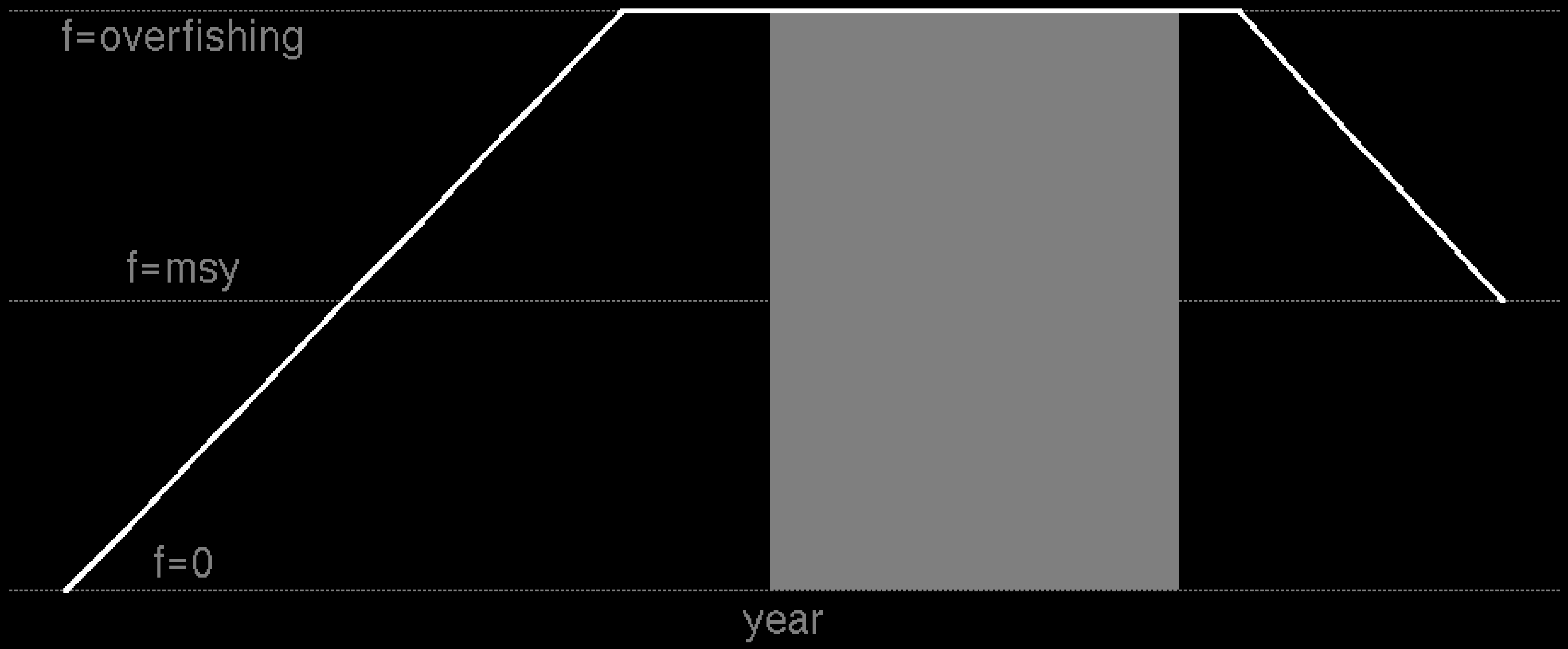
Algorithm step 02 - simulate exploitation



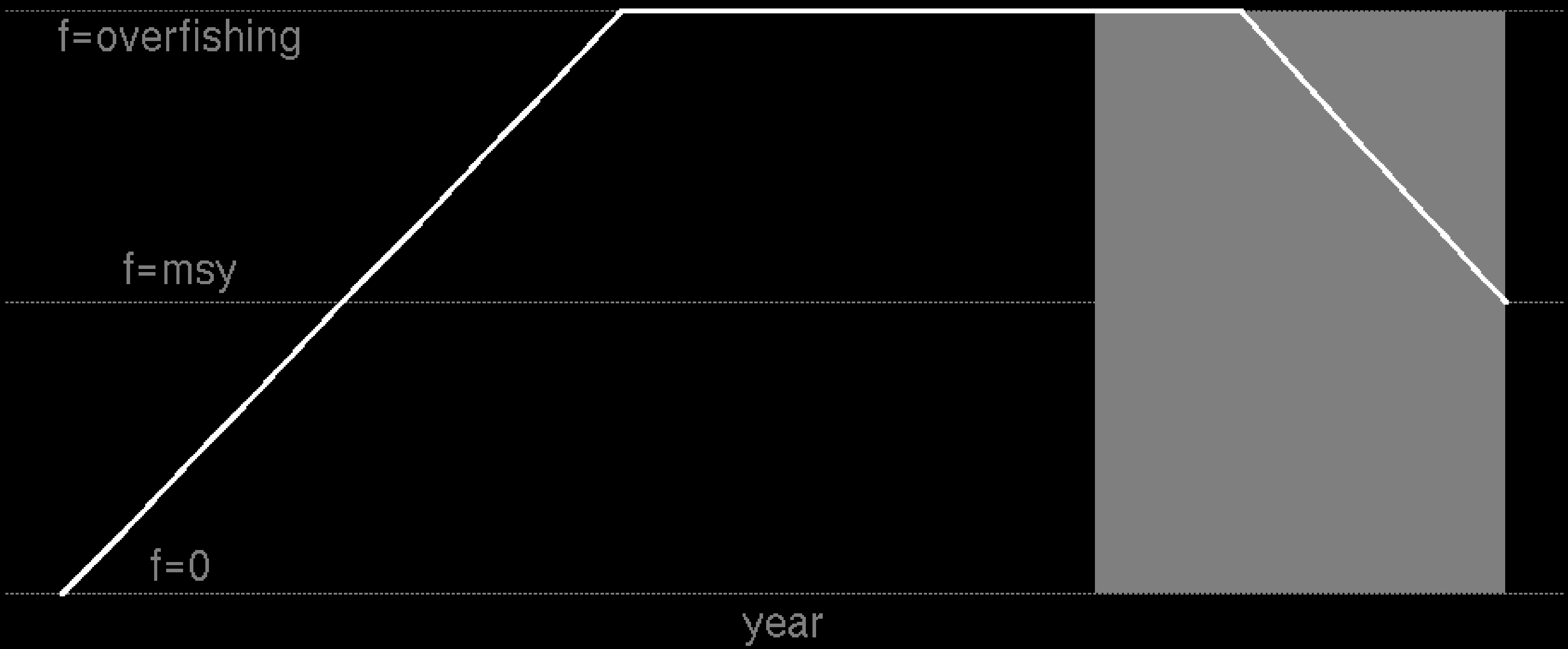
Algorithm step 02 - simulate exploitation



Algorithm step 02 - simulate exploitation



Algorithm step 02 - simulate exploitation



European Commission

Algorithm step 02 - simulate exploitation

(0) full series

(i) "development"

(ii) "development plus over-exploitation"

(iii) "over-exploitation"

(iv) "recovery"

Algorithm step 02 - simulate exploitation

The exploitation pattern was:

- *Full exploited age: 0.7 or 1 * a50*
- *Shape: flat, double normal, “logistic”*

European Commission

Algorithm step 03 - add observation error

- *in abundance indices*
  catchability constant or increase 5% year
  independent lognormal errors cv = 0.2 or 0.5

- *in catch in numbers at age*
  independent lognormal errors cv = 0.1 or 0.3

Algorithm step 04 – fit assessment models

A total of 30 assessment models were built by combining 3 distinct fisheries models, 5 distinct catchability models and 2 distinct stock recruitment models.

For each simulation one combination was randomly chosen to be used in the model.

Algorithm step 04 – fit assessment models

submodel	code	formula
fishery	fm1	~factor(age) + factor(year)
fishery	fm2	~bs(age, 4) + bs(year, 10)
fishery	fm3	~te(age, year, bs = c("tp", "tp"), k = c(4, 15))
catchability	qm0	~1
catchability	qm1	~age
catchability	qm2	~factor(age)
catchability	qm3	~bs(age, 4)
catchability	qm4	~bs(age, 4) + bs(year, 15)
recruitment	rm1	~factor(year)
recruitment	rm2	~bs(year, 15)

Algorithm step 05 – compute statistics

Relative bias and mean square error

SSB, F, C, q, R

Finally:

Scenarios = 224

Species = 1053

Exploitation trajectories = 5

Total runs = 1.15 million

European Commission

Results:

At this point it was clear we couldn't analyse the results in a conventional way.

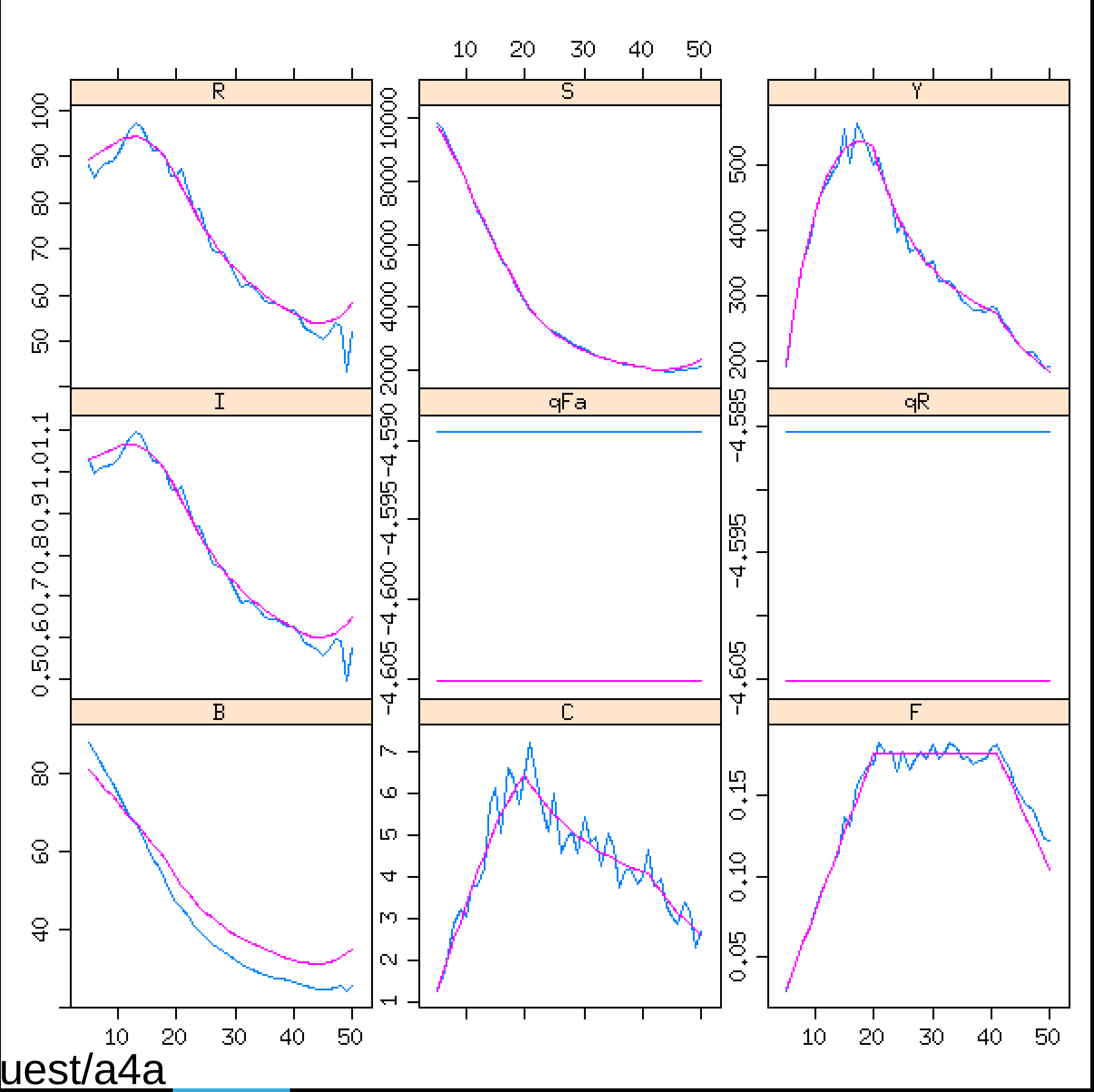
Website:

The website stores, shows and shares the results of the model tests.

https://fishreg.jrc.ec.europa.eu/web/a4a/simulation-testing

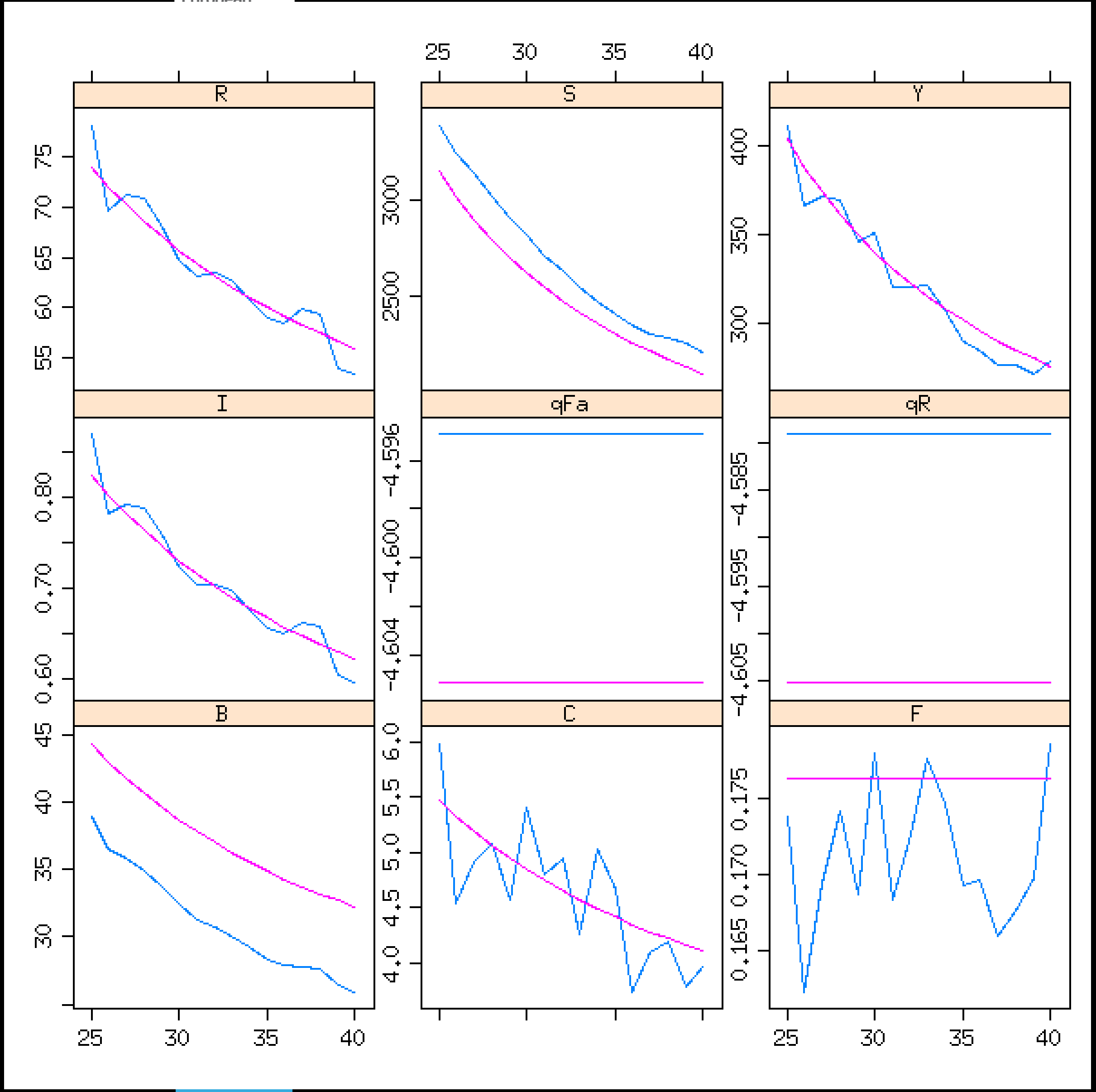
Testing, 1,2 …

WKLIFE stocks
Fishbase stocks(*)



(*) http://fishnet-dev.jrc.it/web/guest/a4a

Testing, 1,2 ...



European Commission

Conclusions:

The simulations allowed us to test

- ➔ *The model capacity to replicate the underlying trends.*
- ➔ *The “automatic mode”.*
- ➔ *“Publish” the results.*

European Commission

Further discussion for the WK.

What do we want to test ?

Rebuild the underlying processes ?

Predict next year catches ?

Inform a harvest control rule ?

Further discussion for the WK.

The objective of the simulation drives the simulation design, methods, performance statistics, etc. Are there particular methods, stats, etc that are more suitable for specific objectives ?

Further discussion for the WK.

The objective of the simulation drives the simulation design, methods, performance statistics, etc. Are there particular simulation designs, performance stats, etc, that are more suitable for specific objectives ?

xxx

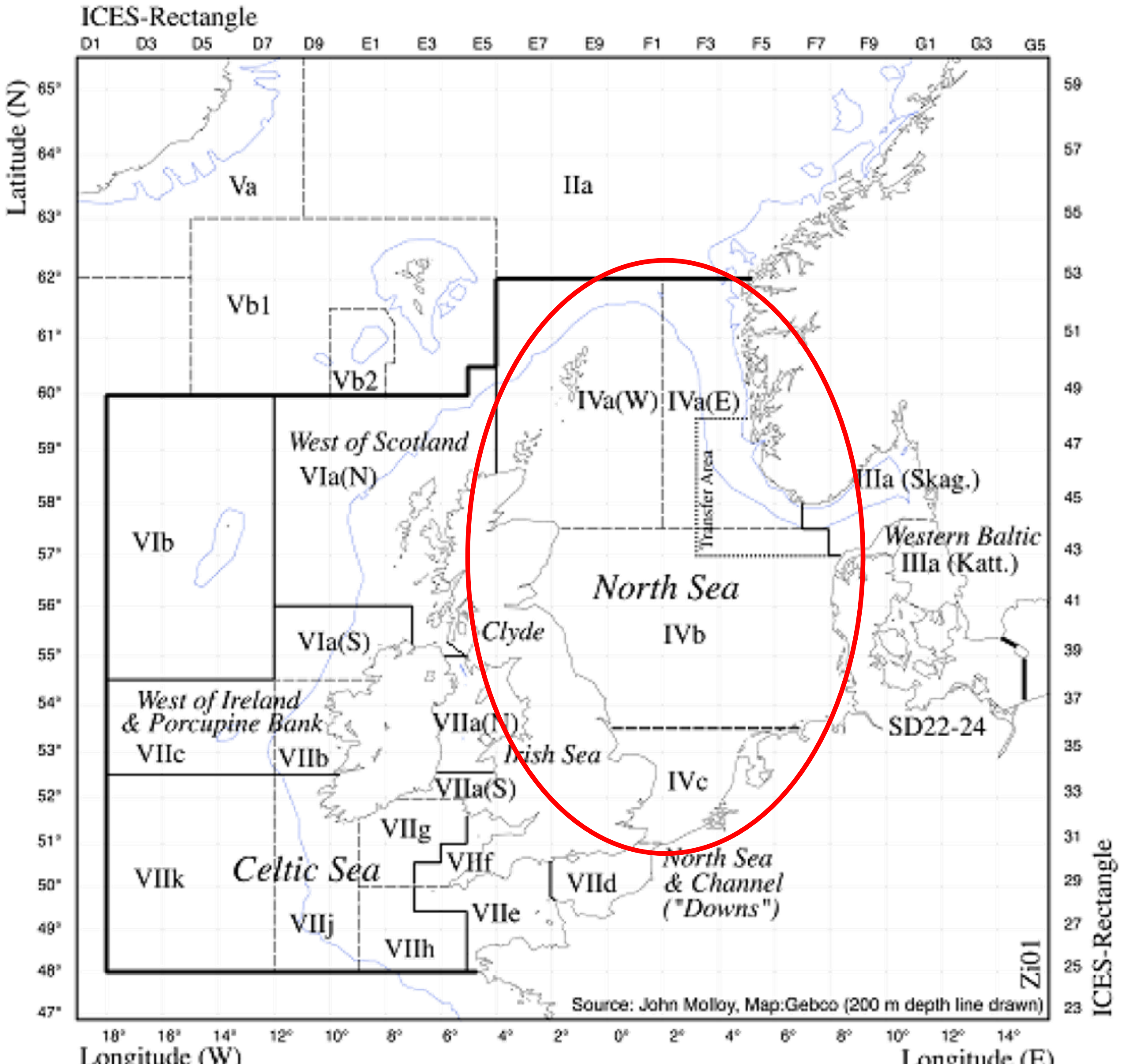
North Sea herring: fits to real data

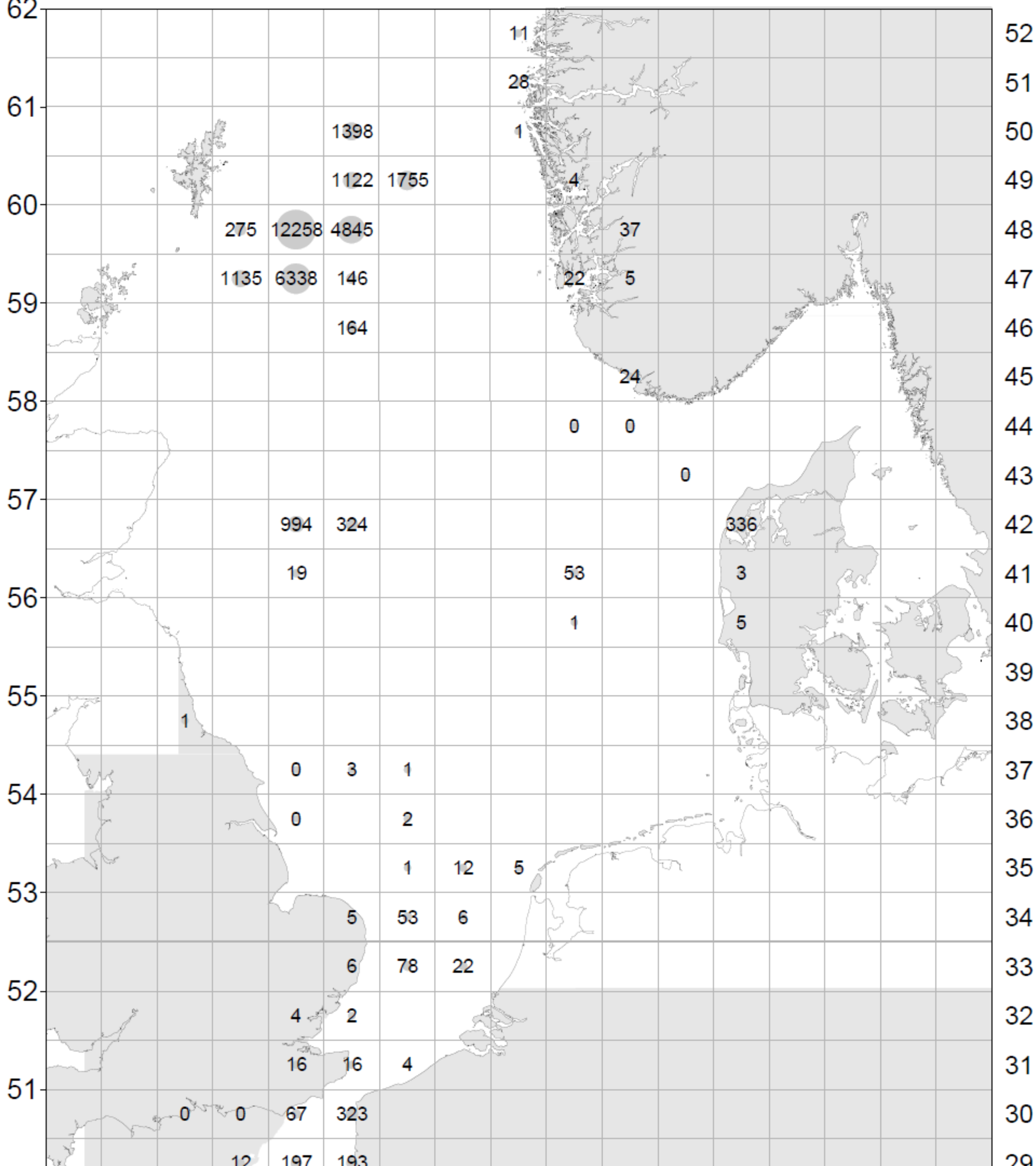
July 2013

Boston MA

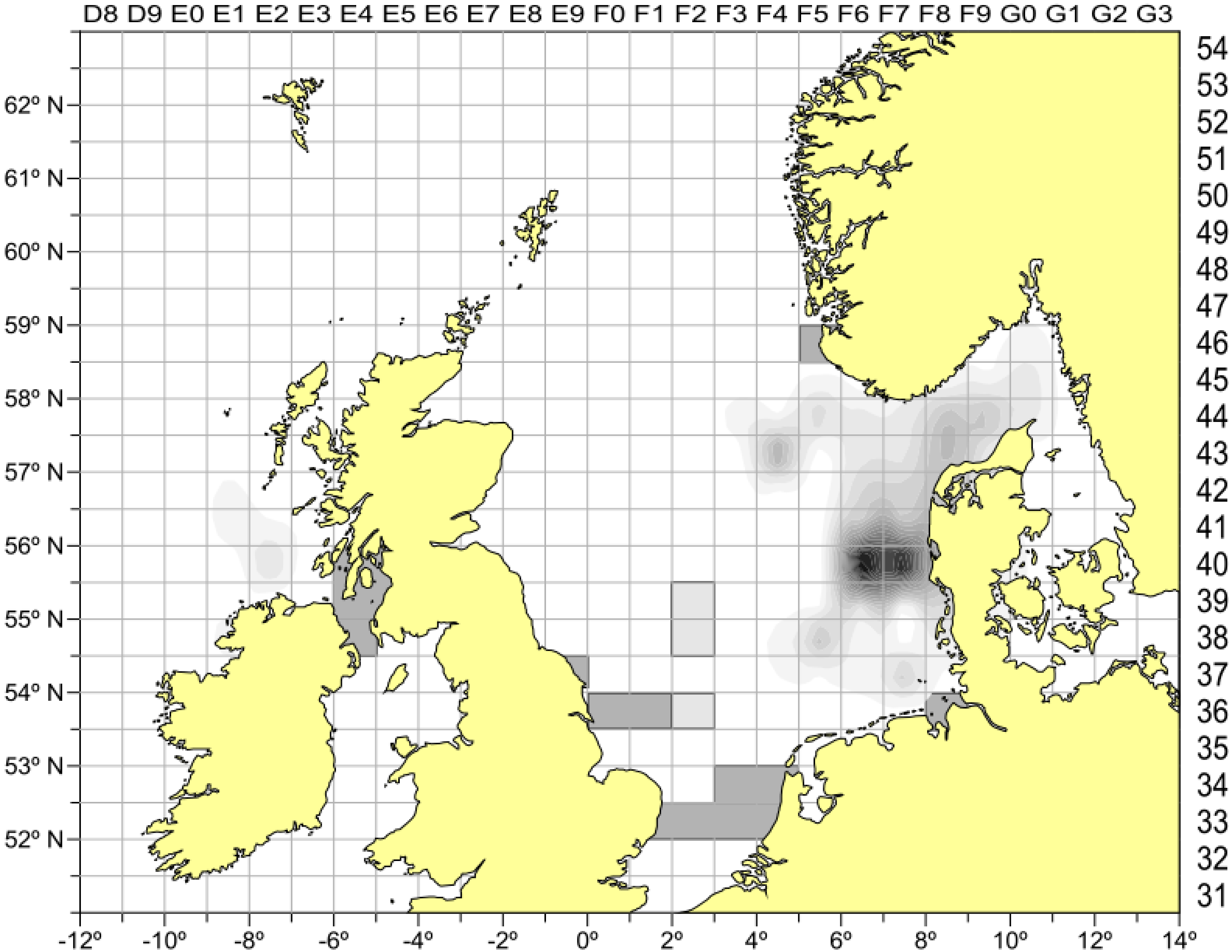
Herring management areas

Gear:

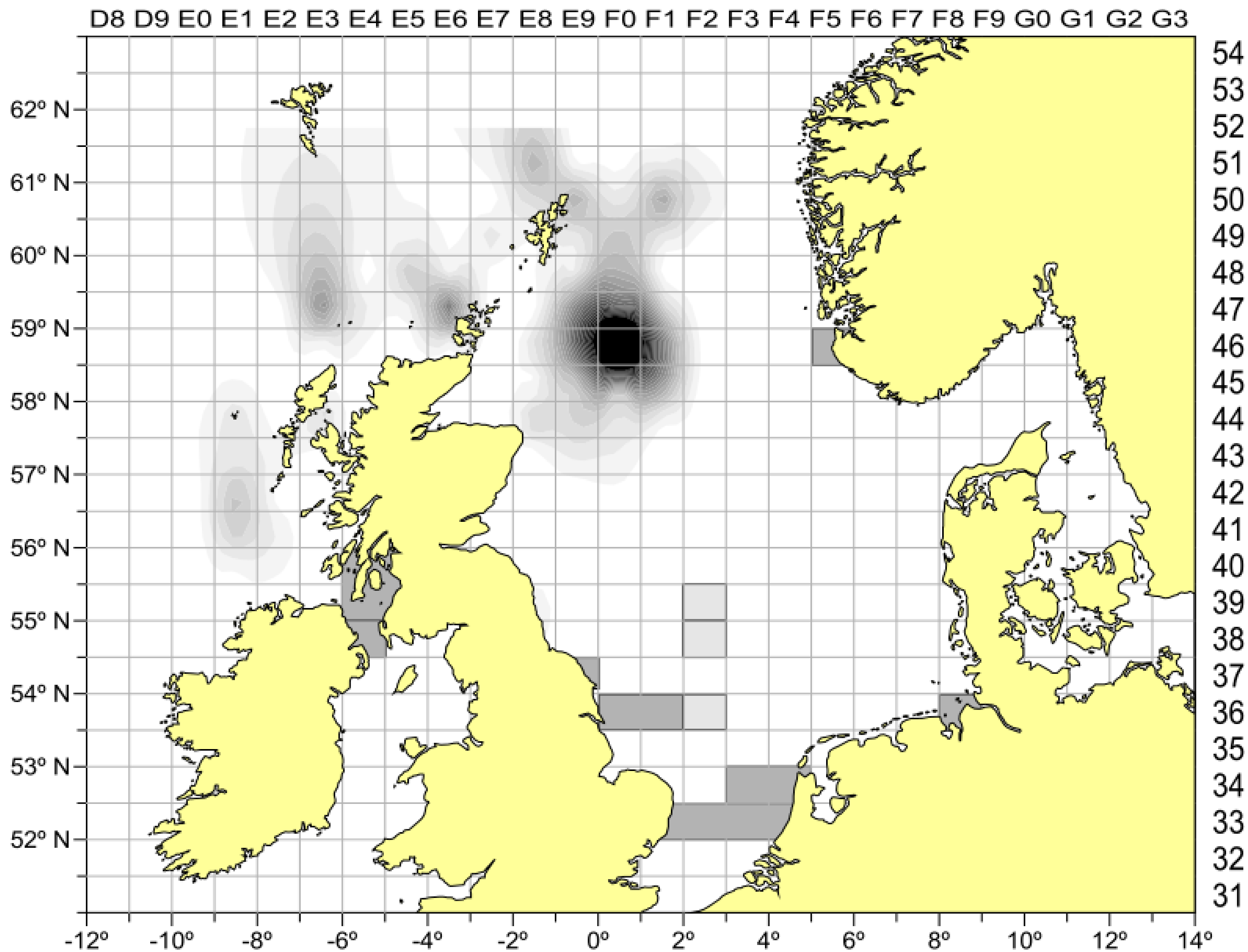



62
61
60
59
58
57
56
55
54
53
52
51
52
51
50
49
48
47
46
45
44
43
42
41
40
39
38
37
36
35
34
33
32
31
30
29
11
28
1
1398
1122
1755
4
275
12258
4845
37
1135
6338
146
22
5
164
24
0
0
0
994
324
336
19
53
3
1
5
1
0
3
1
0
2
1
12
5
5
53
6
6
78
22
4
2
16
16
4
0
0
67
323
12
197
193

2011 Immature June-July survey



2011 **Mature** June-July survey



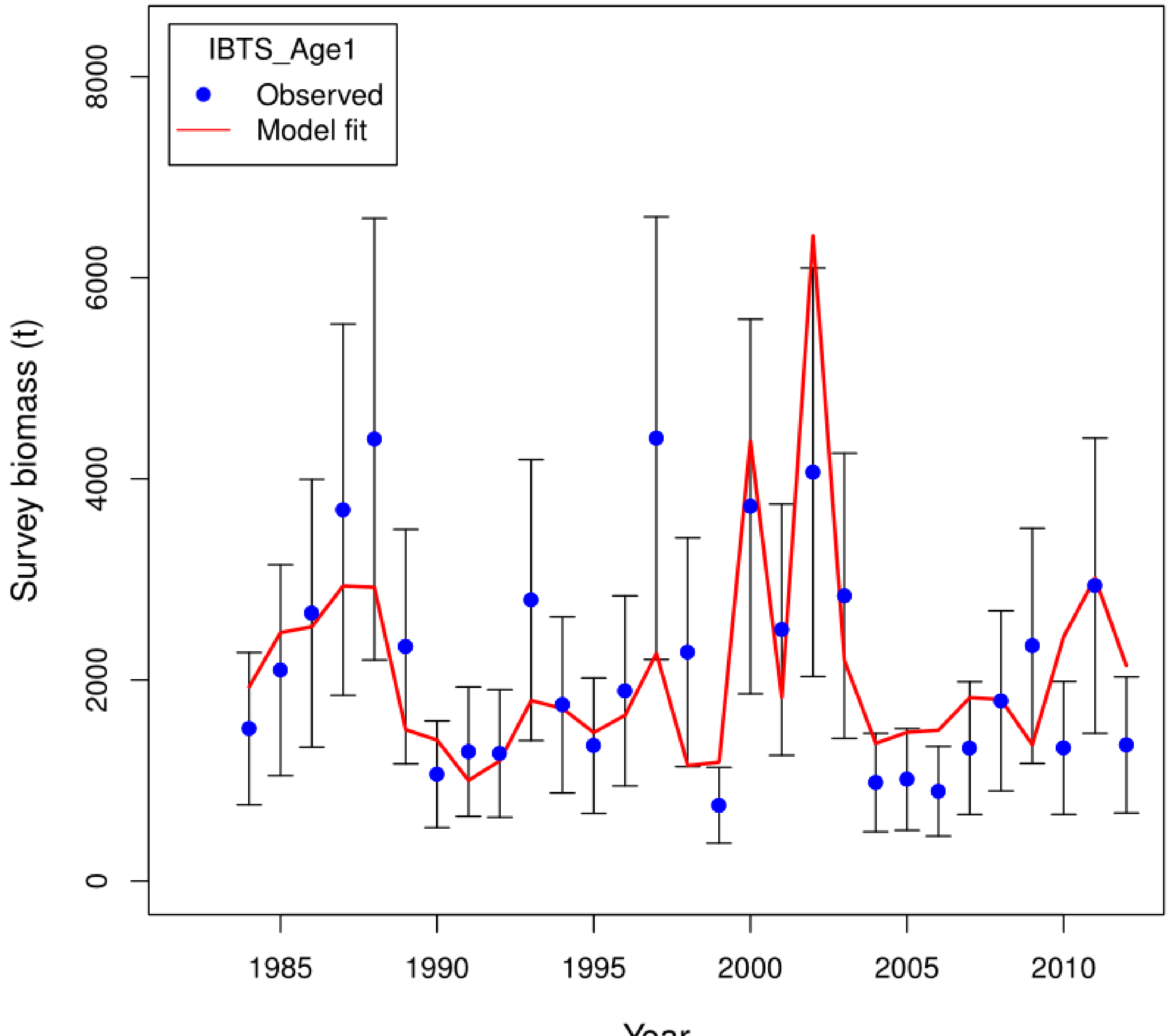
NS Herring distributed data

- My experience
  - Ignored HERAS data prior to 1997 because age 0s absent
  - Only used ITBS recruitment index
- My runs included:
  - Model 1: constant selectivity
  - Model 2: time-varying selectivity

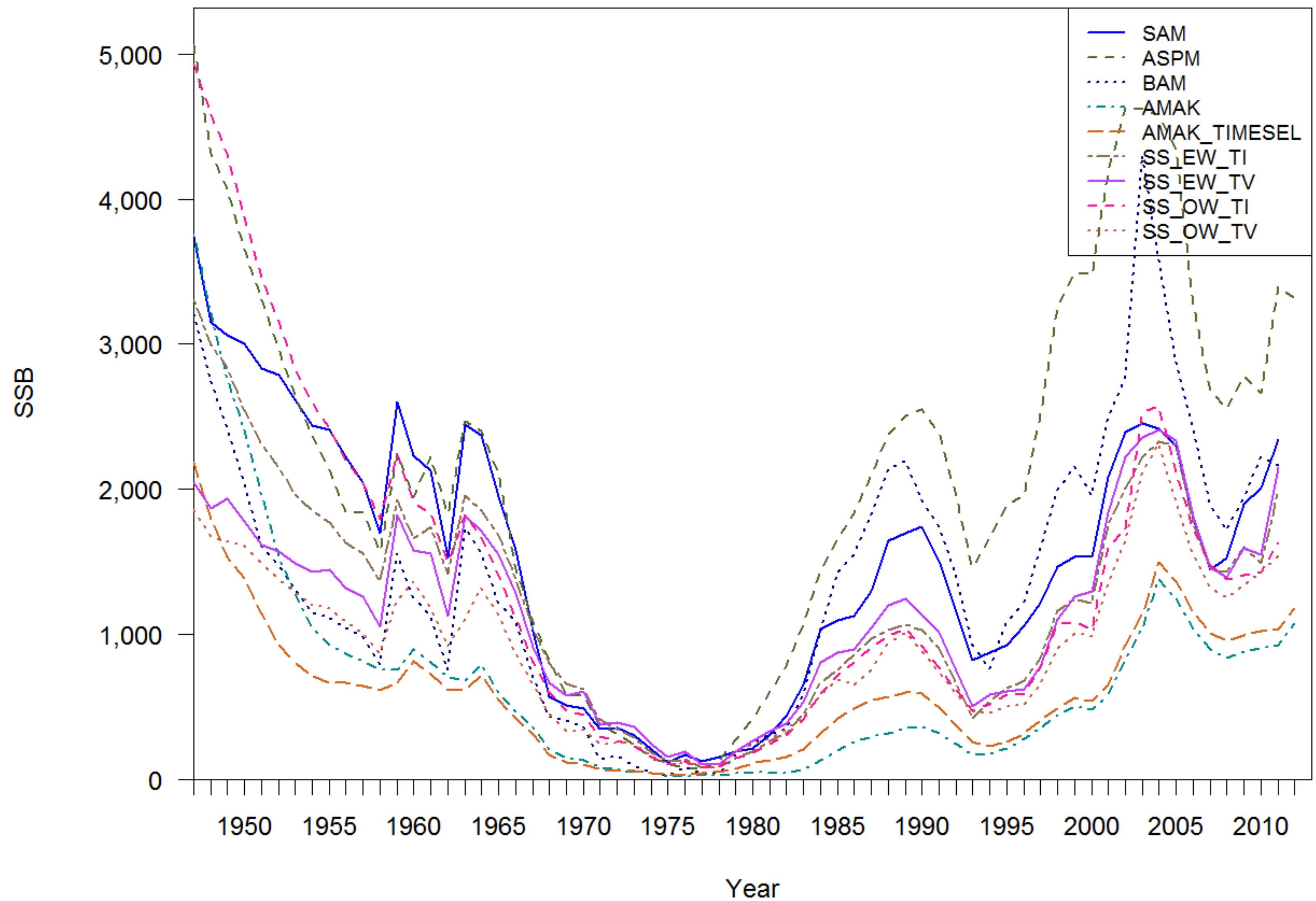
Data

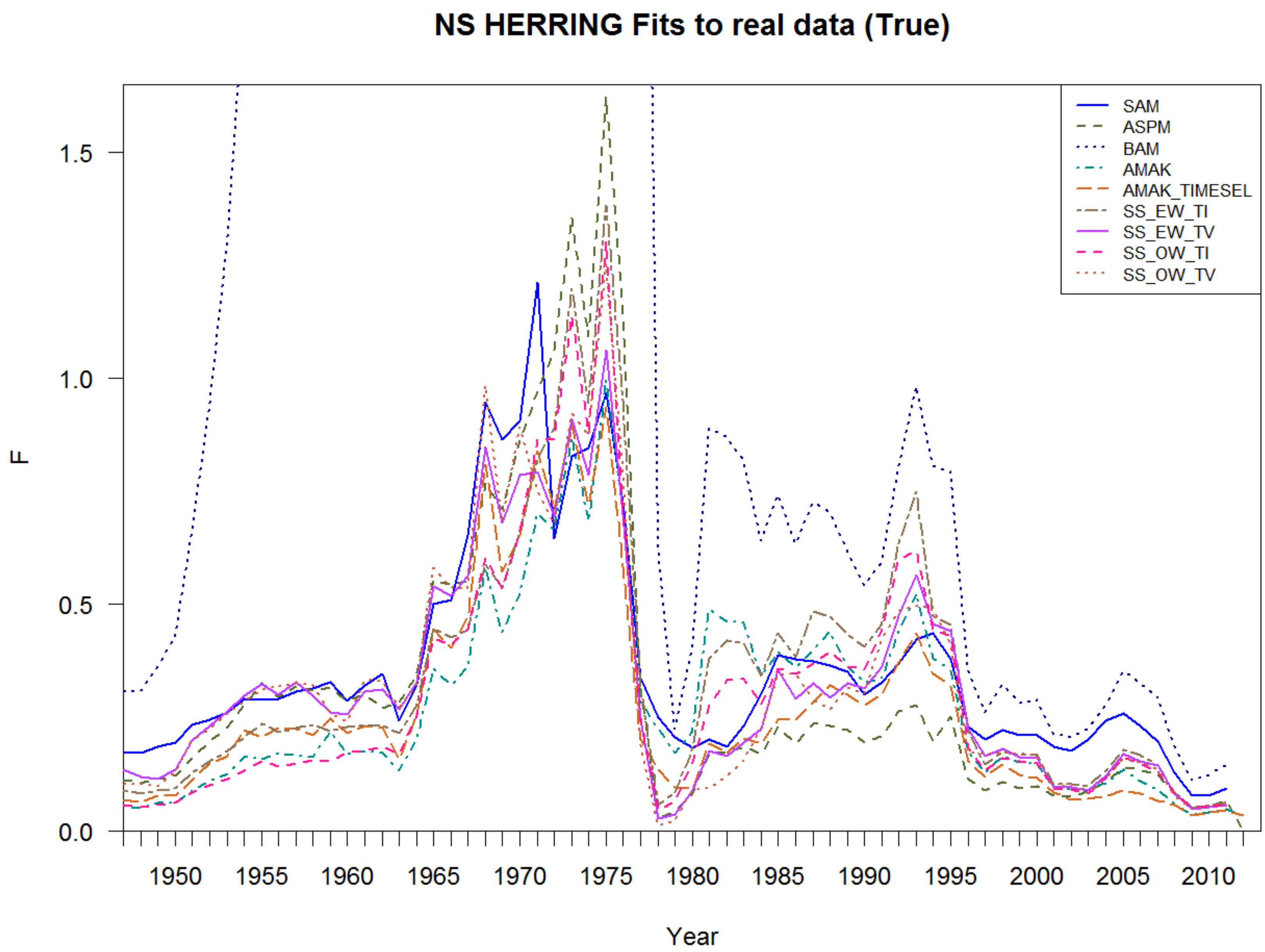
- Catch 1947-2012
- IBTS age 1 1984-2011
- Spawning index 1970-2012
- HERAS 1995-2012

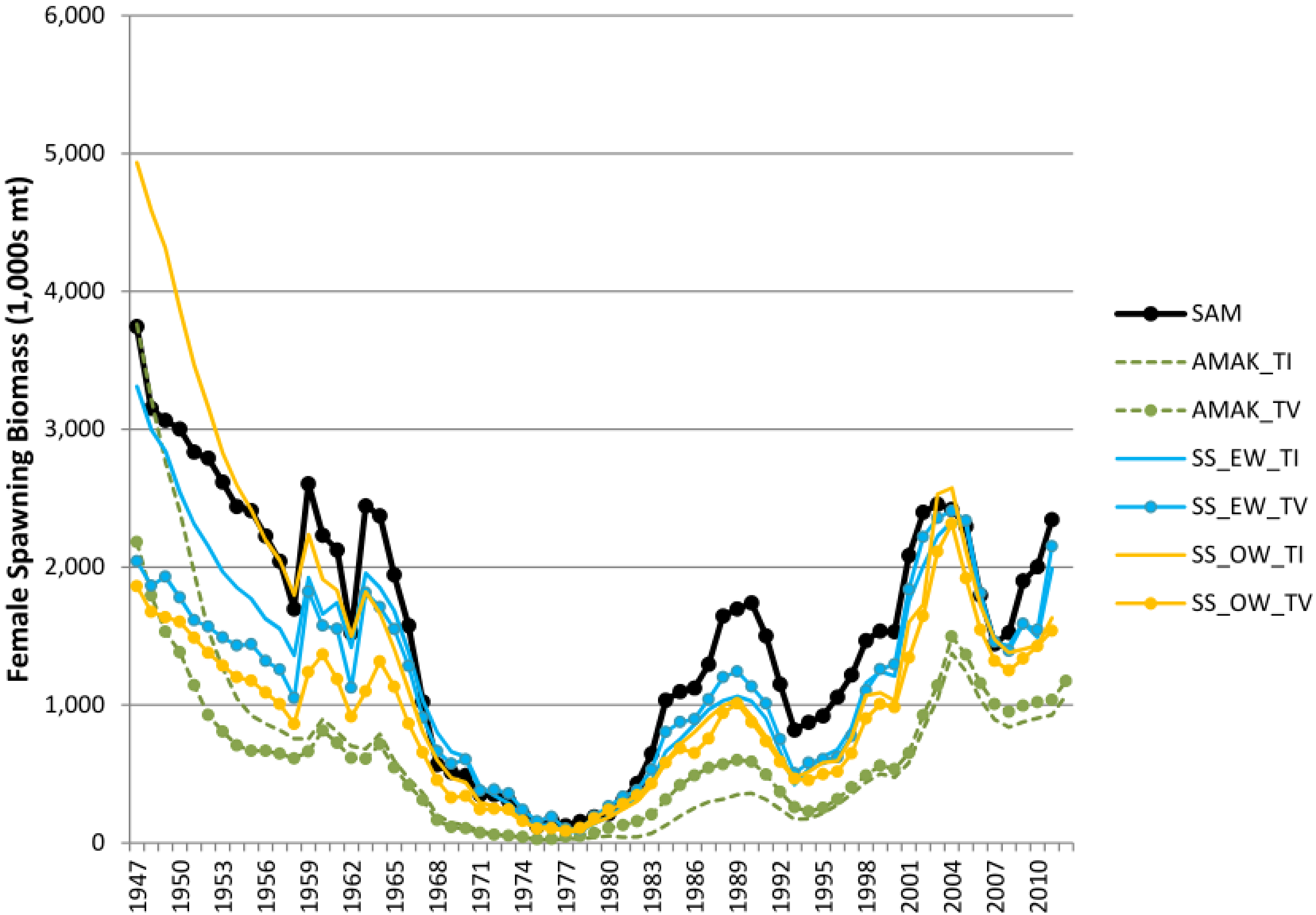
Age-1 survey (ITBS)



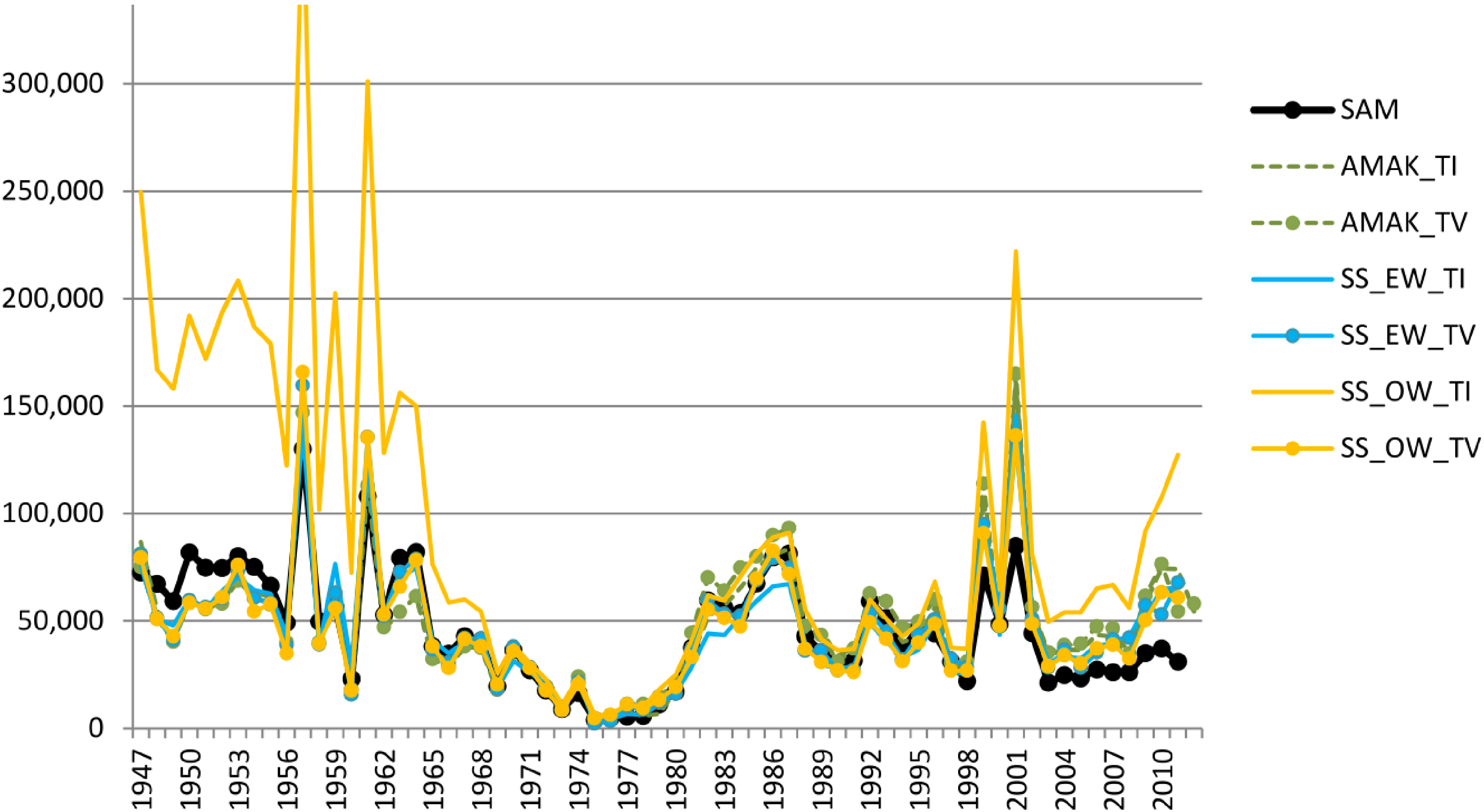
NS HERRING Fits to real data (True)



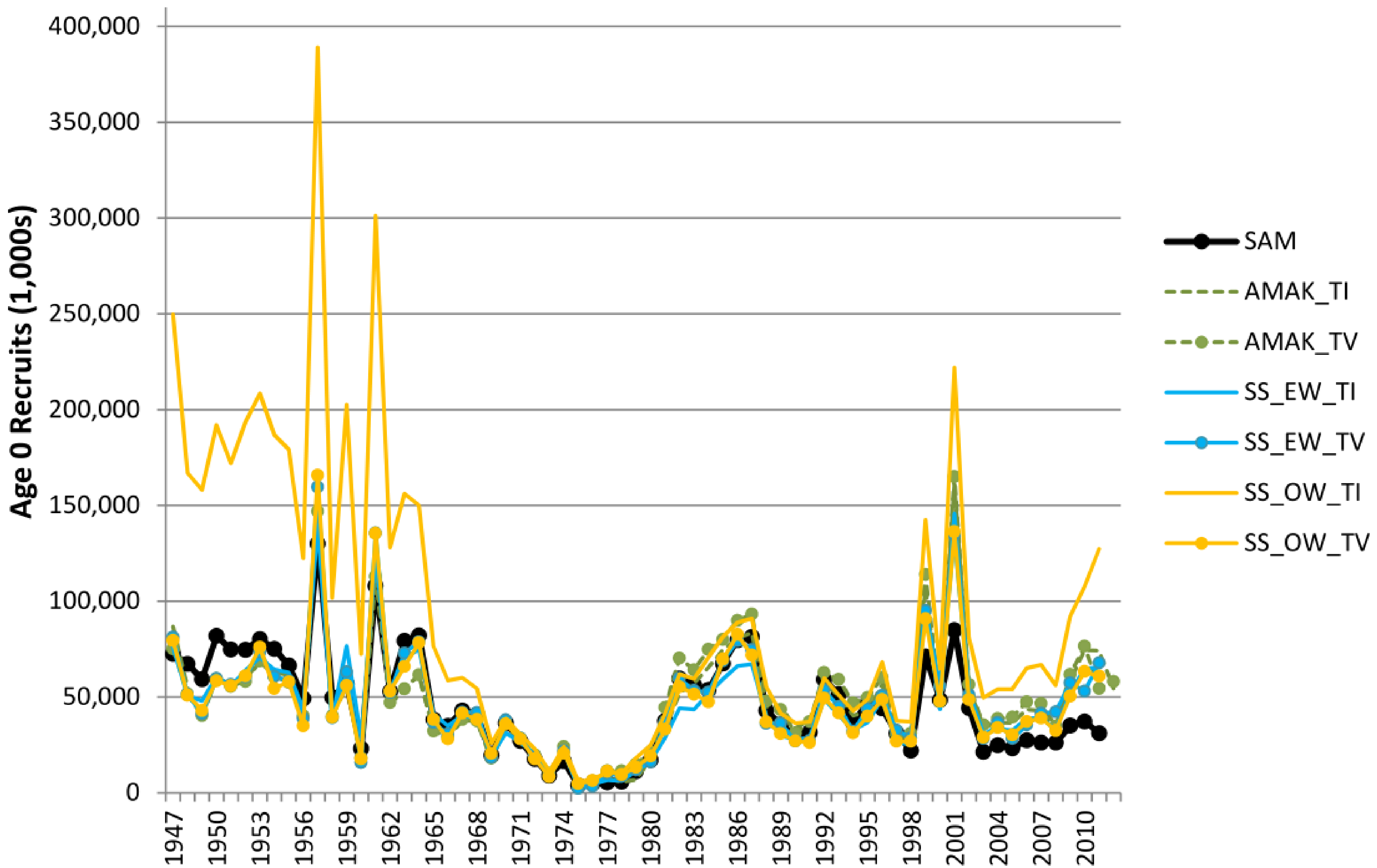

NS HERRING Fits to real data (True)
SAM
ASPM
BAM
AMAK
AMAK_TIMESEL
SS_EW_TI
SS_EW_TV
SS_OW_TI
SS_OW_TV
F
0.0
0.5
1.0
1.5
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
Year



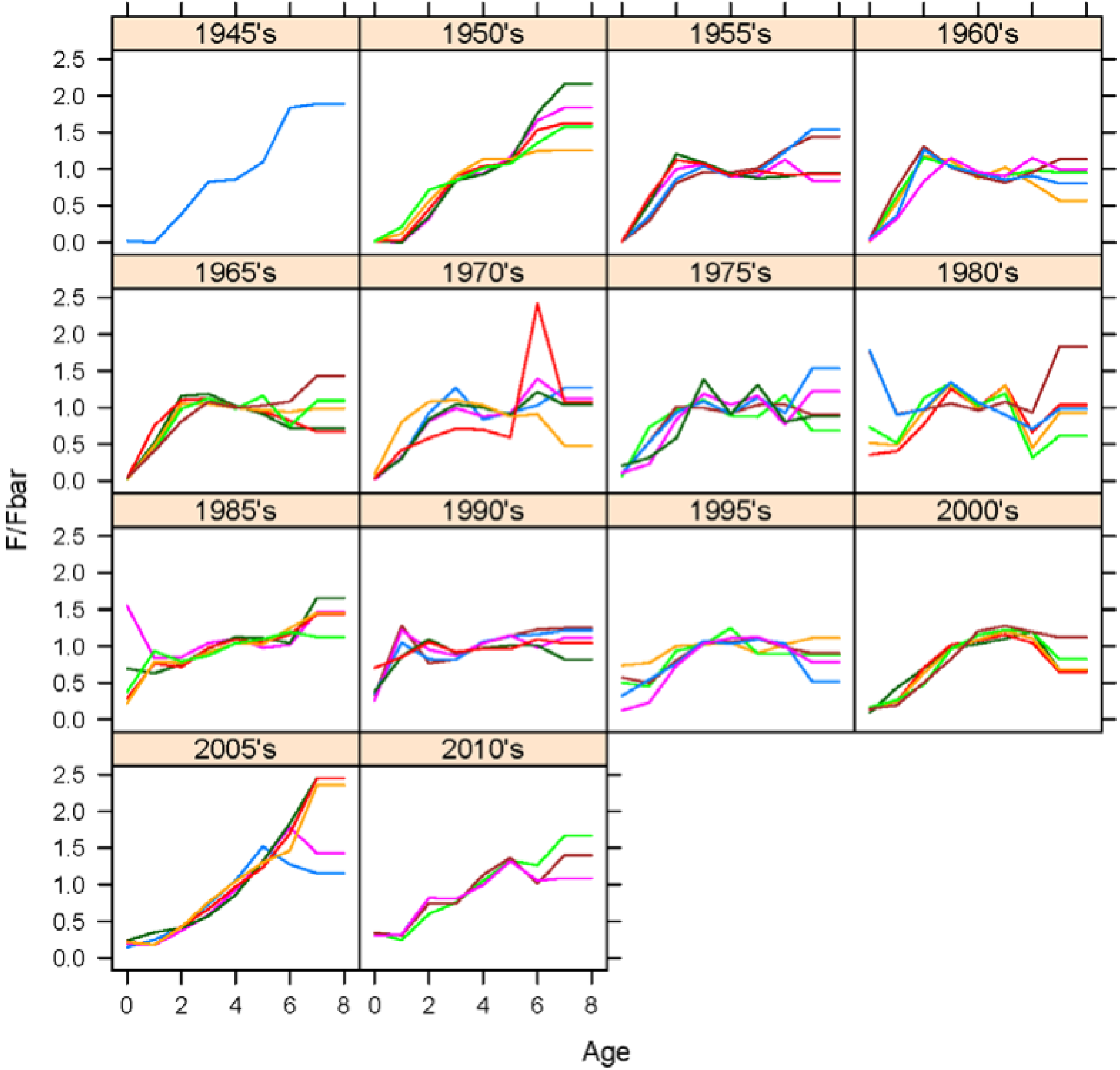
Female Spawning Biomass (1,000s mt)
6,000
5,000
4,000
3,000
2,000
1,000
0
1947
1950
1953
1956
1959
1962
1965
1968
1971
1974
1977
1980
1983
1986
1989
1992
1995
1998
2001
2004
2007
2010
SAM
AMAK_TI
AMAK_TV
SS_EW_TI
SS_EW_TV
SS_OW_TI
SS_OW_TV

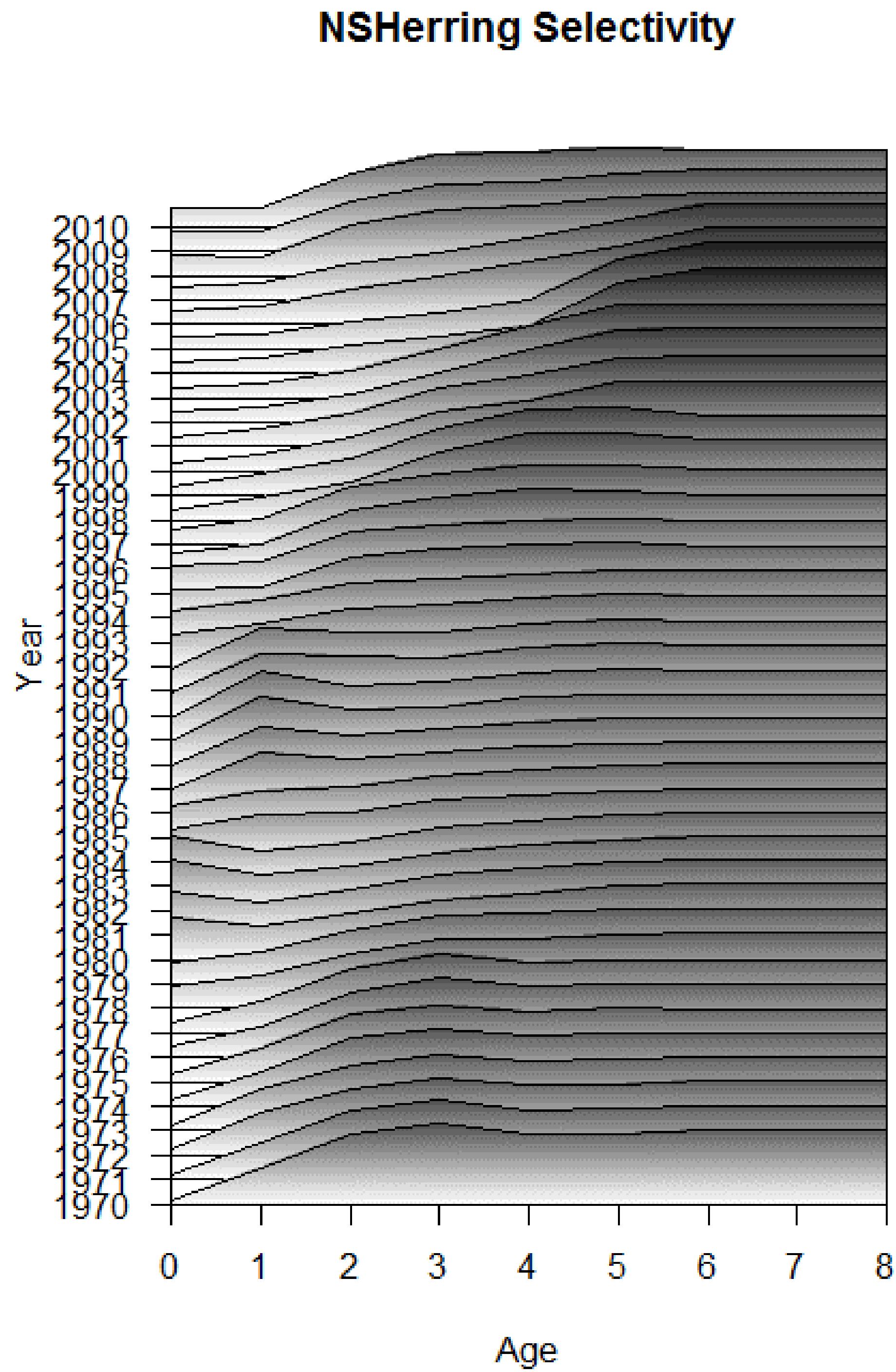


300,000
250,000
200,000
150,000
100,000
50,000
0
1947
1950
1953
1956
1959
1962
1965
1968
1971
1974
1977
1980
1983
1986
1989
1992
1995
1998
2001
2004
2007
2010
SAM
AMAK_TI
AMAK_TV
SS_EW_TI
SS_EW_TV
SS_OW_TI
SS_OW_TV

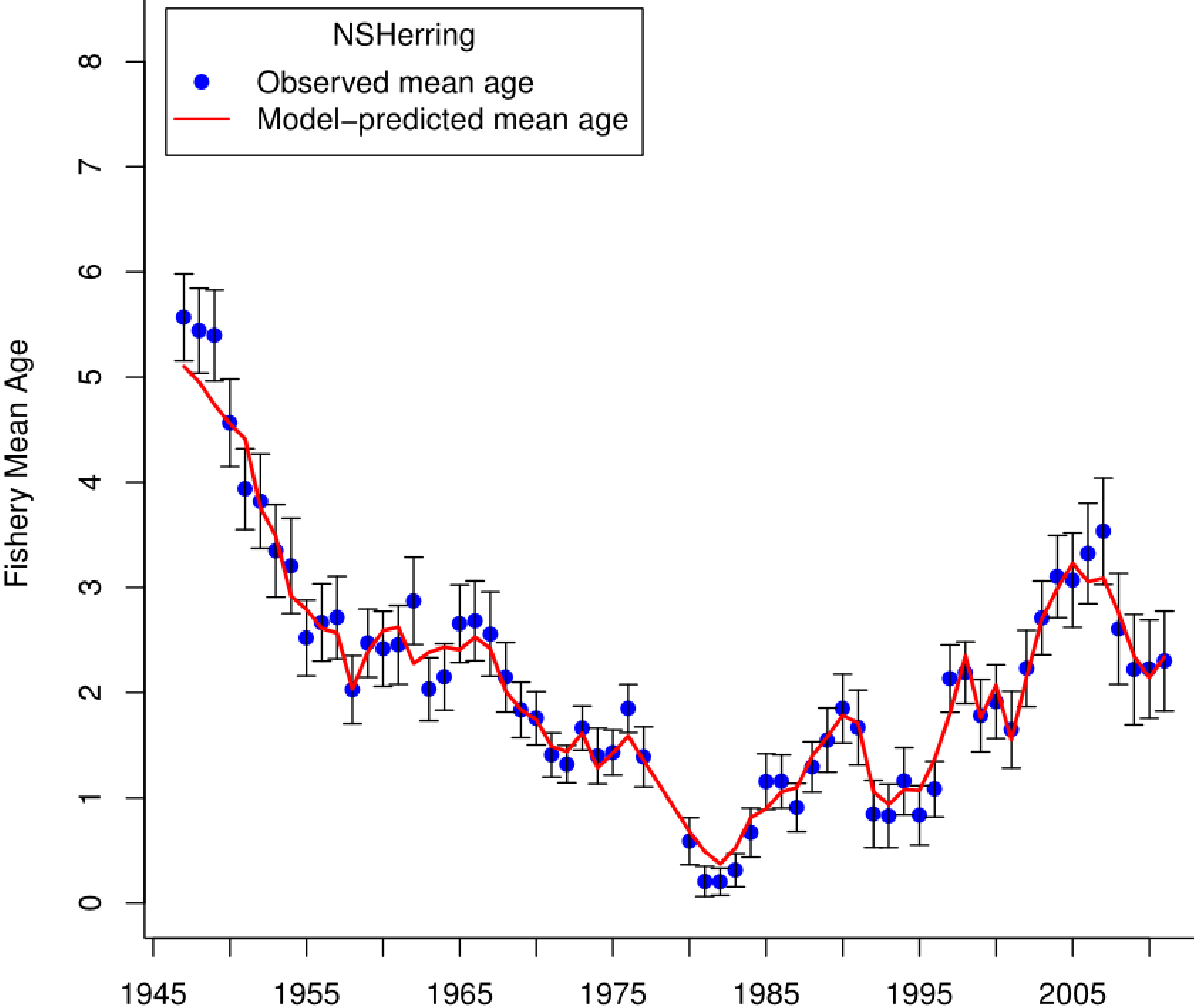


Age 0 Recruits (1,000s)
400,000
350,000
300,000
250,000
200,000
150,000
100,000
50,000
0
1947
1950
1953
1956
1959
1962
1965
1968
1971
1974
1977
1980
1983
1986
1989
1992
1995
1998
2001
2004
2007
2010
SAM
AMAK_TI
AMAK_TV
SS_EW_TI
SS_EW_TV
SS_OW_TI
SS_OW_TV


1945's
1950's
1955's
1960's
1965's
1970's
1975's
1980's
1985's
1990's
1995's
2000's
2005's
2010's
F/Fbar
Age
0.0
0.5
1.0
1.5
2.0
2.5
0
2
4
6
8



NSHerring Selectivity
Year
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
Age
0
1
2
3
4
5
6
7
8


NSHerring
Observed mean age
Model−predicted mean age
Fishery Mean Age
0
1
2
3
4
5
6
7
8
1945
1955
1965
1975
1985
1995
2005

Discussion points

Haddock

Anders Nielsen
an@aqua.dtu.dk

DTU Aqua
National Institute of Aquatic Resources

Haddock

Data

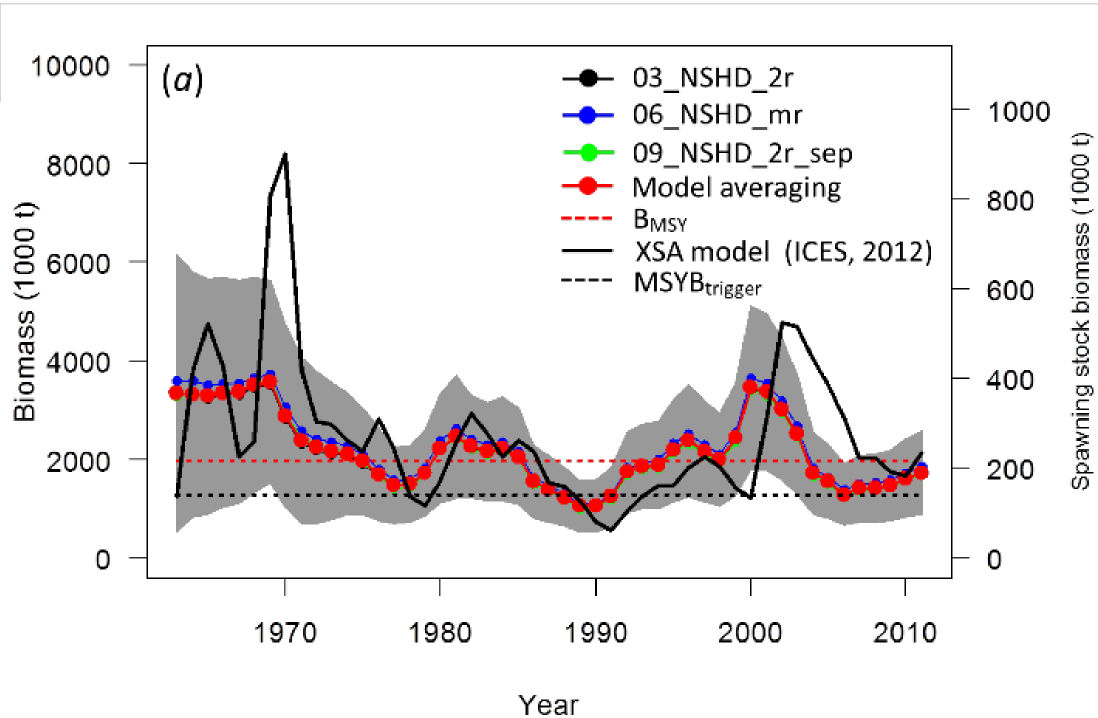
- Catch at age available from 1963
- Five surveys available first one starting in 1977
- Spike recruitment's
- Most models agree (mostly) on main trends

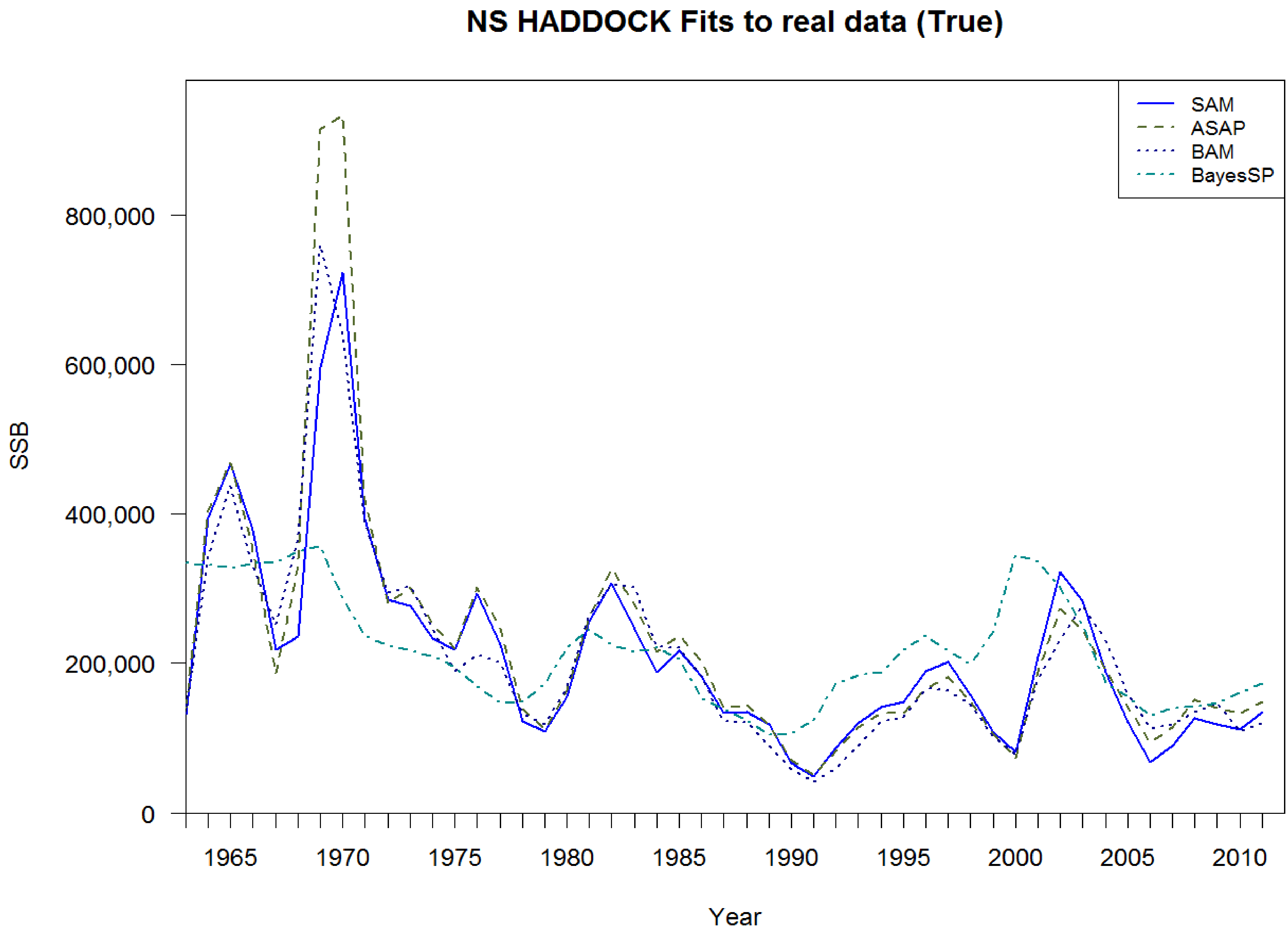
Models

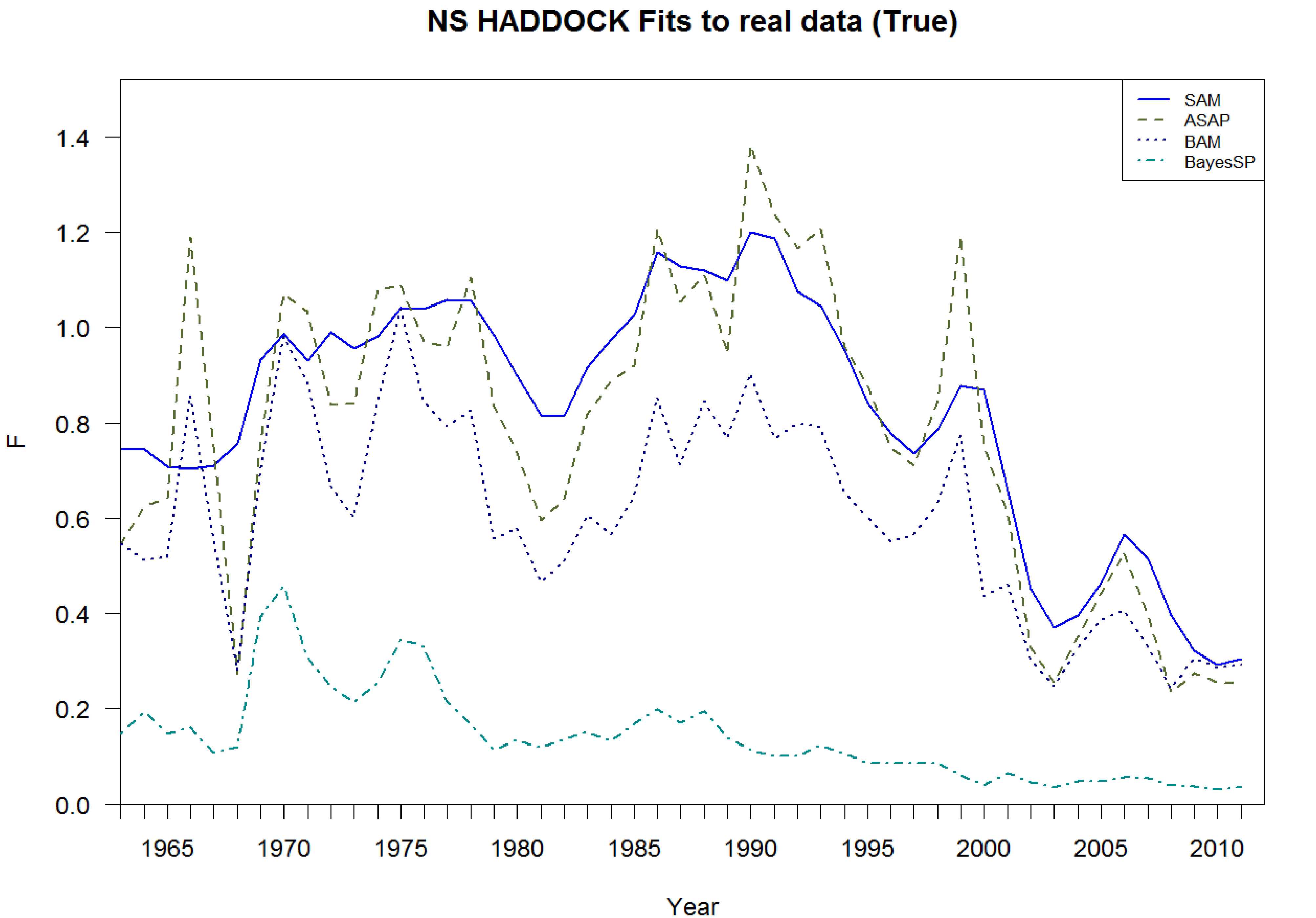
- BAM (results uploaded)
- ASAP (results uploaded)
- BayesSP (results uploaded)
- Adapt (sim study)
- XSA (sim study)
- SAM (results uploaded, sim study)
- Gudmundsson & Gunnlaugsson state-space model (applied from 1992, separate presentation)

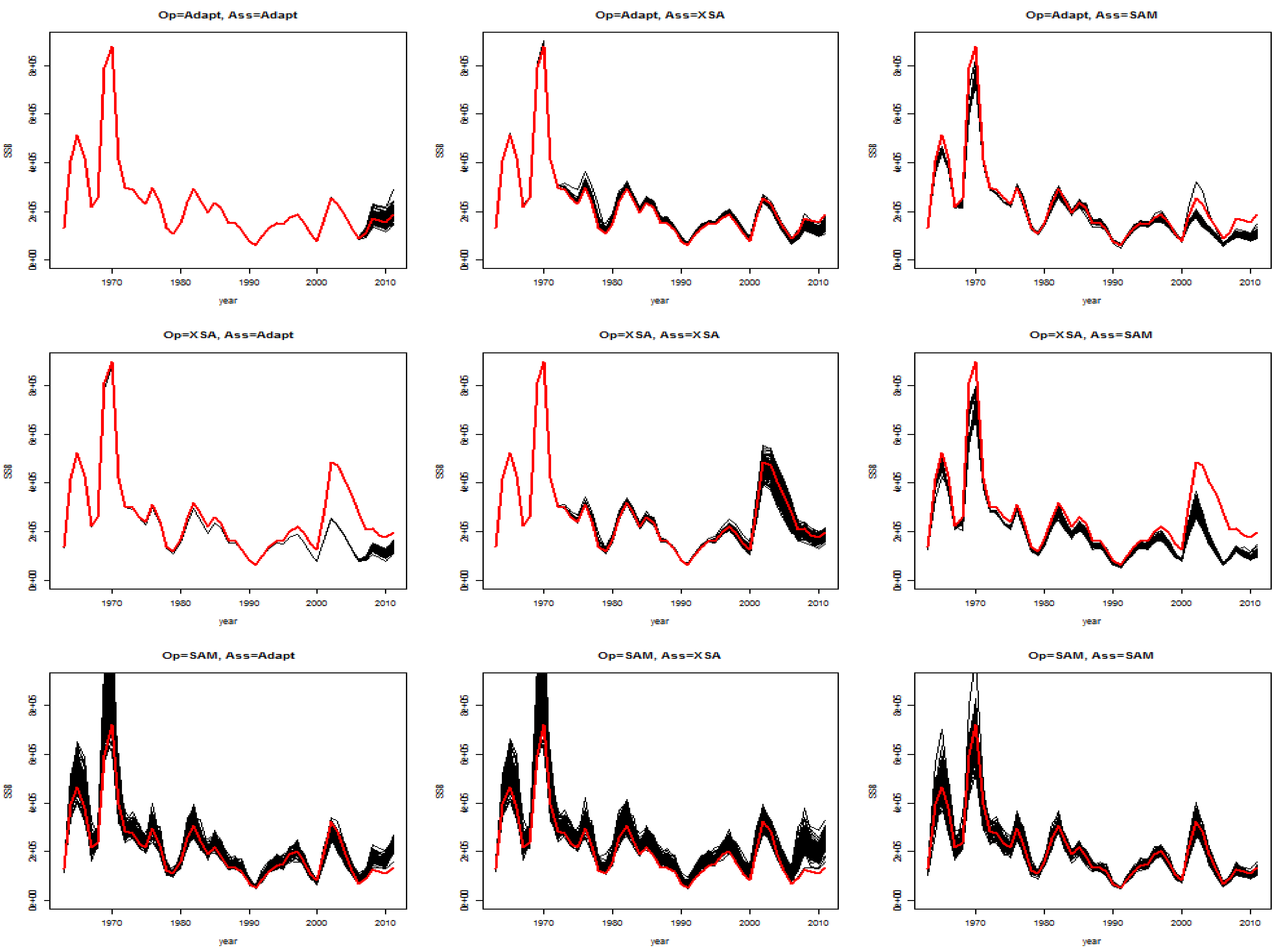
BayesSP

- Hierarchically structured Bayesian surplus production models
- Landings (1963 - 2011)
- Five survey indices (converted to biomass):
- Log-normal prior distributions for intrinsic growth rate and carrying capacity were assumed
- Lot of effort in model selection
- Three models provided the most credible fits
- Model averaging was applied to summarize the results of the three most credible models

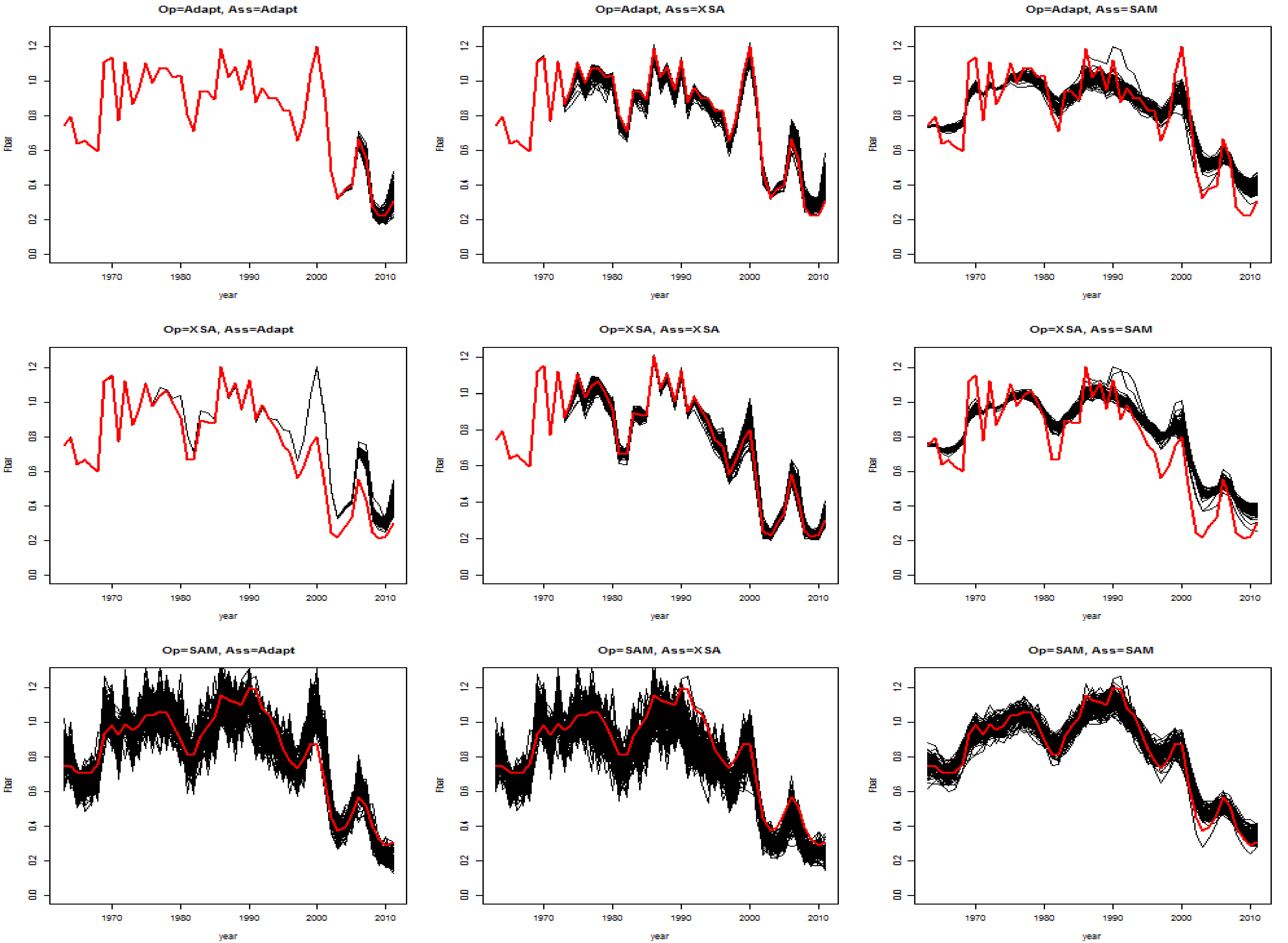



NS HADDOCK Fits to real data (True)
SAM
ASAP
BAM
BayesSP
SSB
800,000
600,000
400,000
200,000
0
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
Year


NS HADDOCK Fits to real data (True)
SAM
ASAP
BAM
BayesSP
F
Year
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010



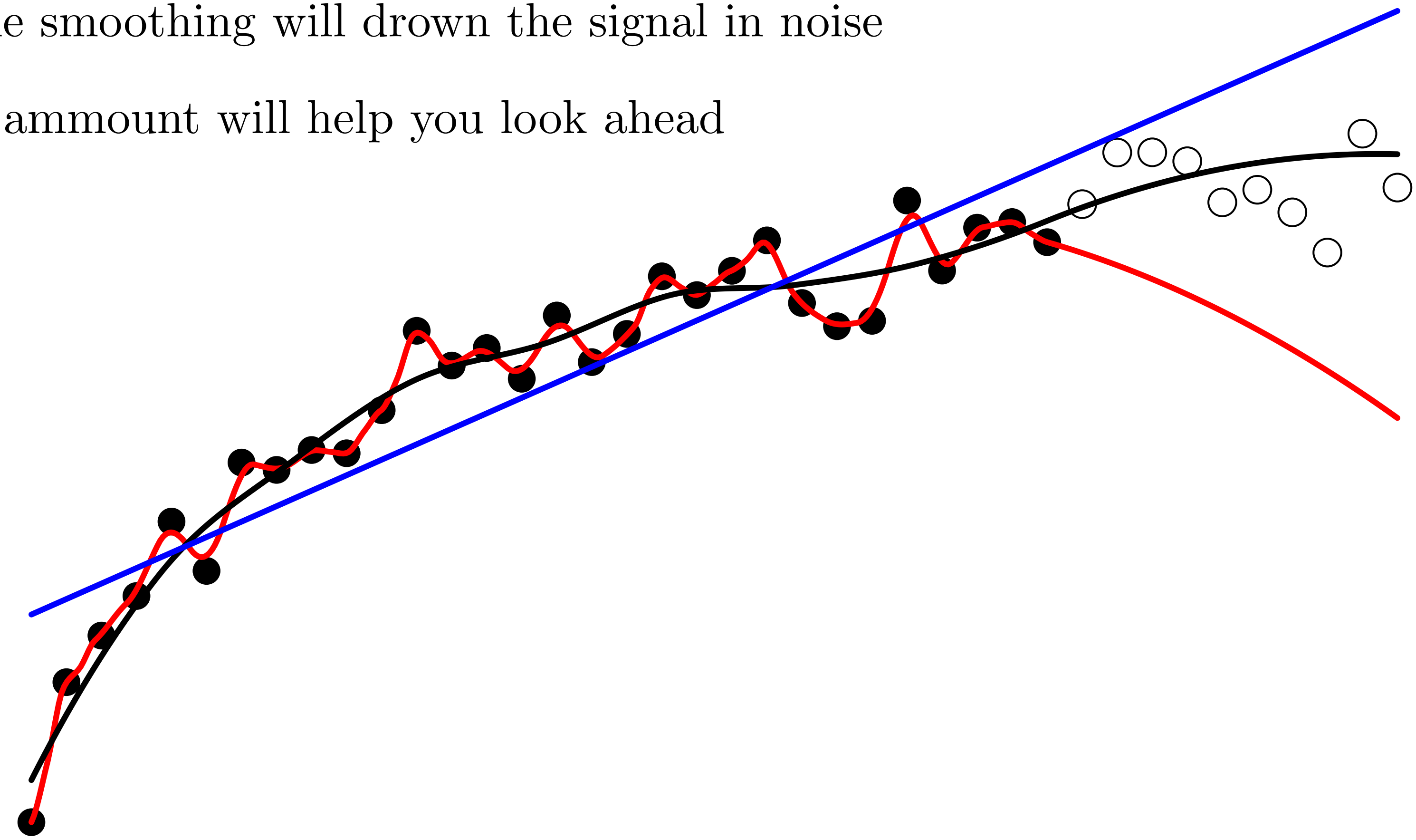
Op=Adapt, Ass=Adapt
Op=Adapt, Ass=XSA
Op=Adapt, Ass=SAM
Op=XSA, Ass=Adapt
Op=XSA, Ass=XSA
Op=XSA, Ass=SAM
Op=SAM, Ass=Adapt
Op=SAM, Ass=XSA
Op=SAM, Ass=SAM
SSB
year
0e+00
2e+05
4e+05
6e+05
8e+05
1970
1980
1990
2000
2010



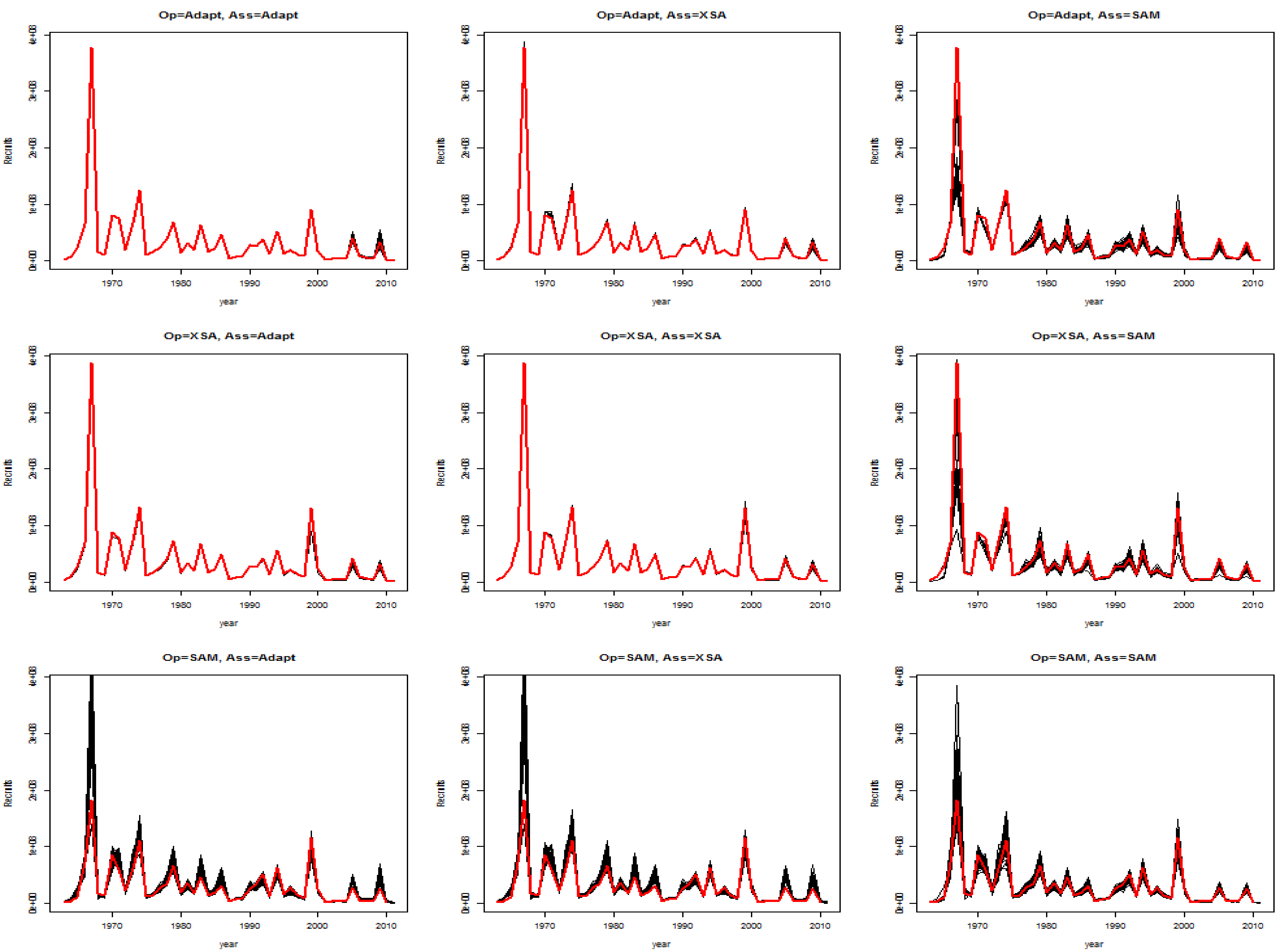
Op=Adapt, Ass=Adapt
Op=Adapt, Ass=XSA
Op=Adapt, Ass=SAM
Op=XSA, Ass=Adapt
Op=XSA, Ass=XSA
Op=XSA, Ass=SAM
Op=SAM, Ass=Adapt
Op=SAM, Ass=XSA
Op=SAM, Ass=SAM
Fbar
year
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1970
1980
1990
2000
2010

Is smoothing evil?

- Too much smoothing will bias the signal
- Too little smoothing will drown the signal in noise
- Correct ammount will help you look ahead

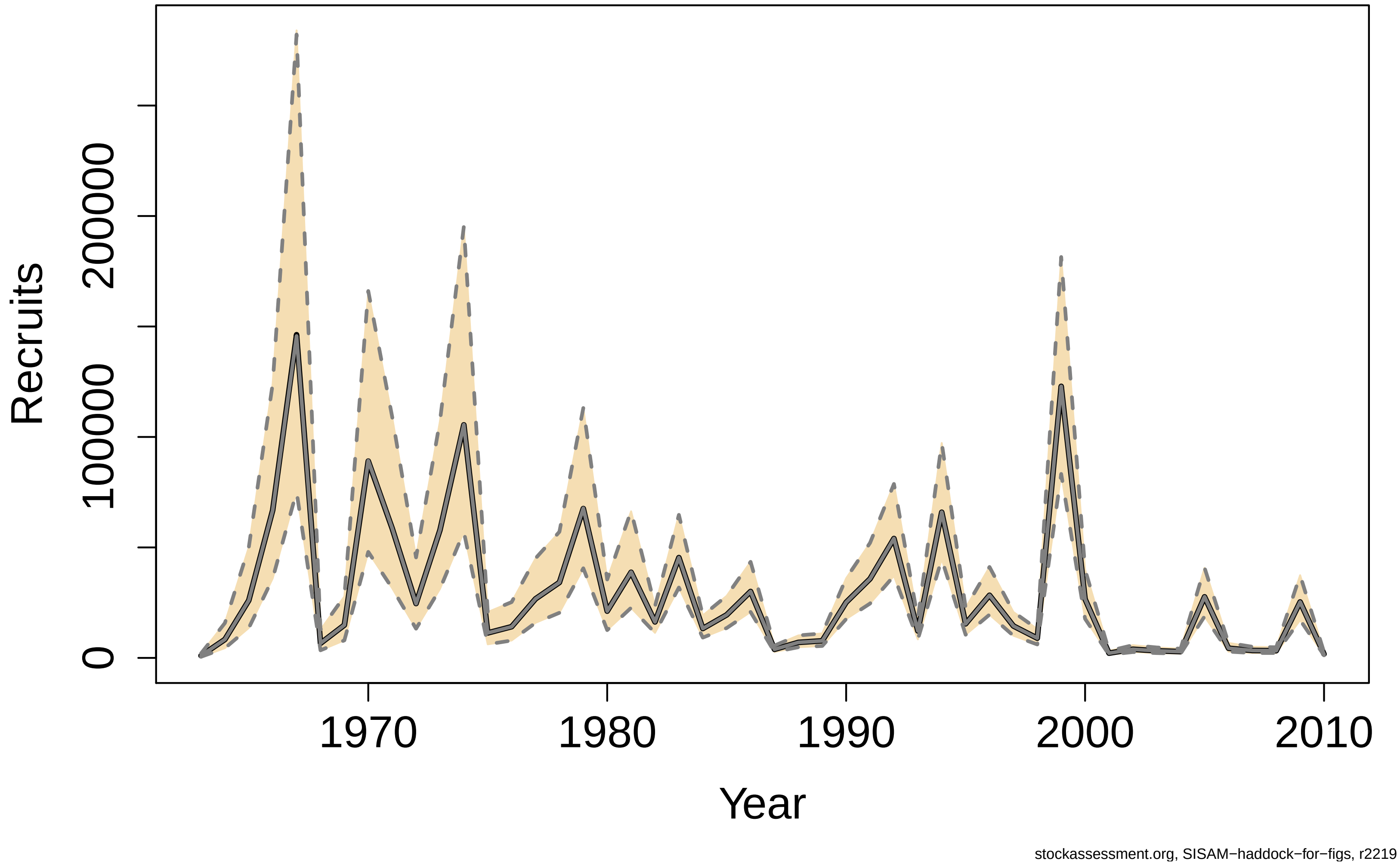


- Correct amount should not be subjective.
- Also, if we keep adding noise to the signal - we should end with all noise and no signal.



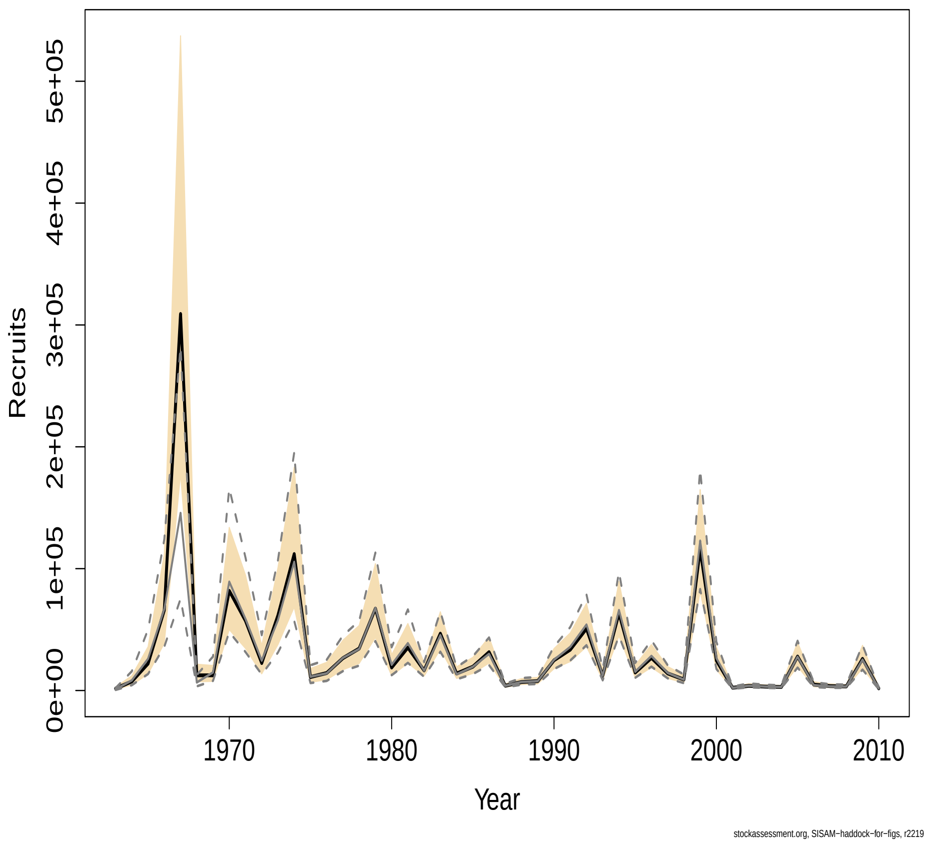
Op=Adapt, Ass=Adapt
Op=Adapt, Ass=XSA
Op=Adapt, Ass=SAM
Op=XSA, Ass=Adapt
Op=XSA, Ass=XSA
Op=XSA, Ass=SAM
Op=SAM, Ass=Adapt
Op=SAM, Ass=XSA
Op=SAM, Ass=SAM
1970
1980
1990
2000
2010
year

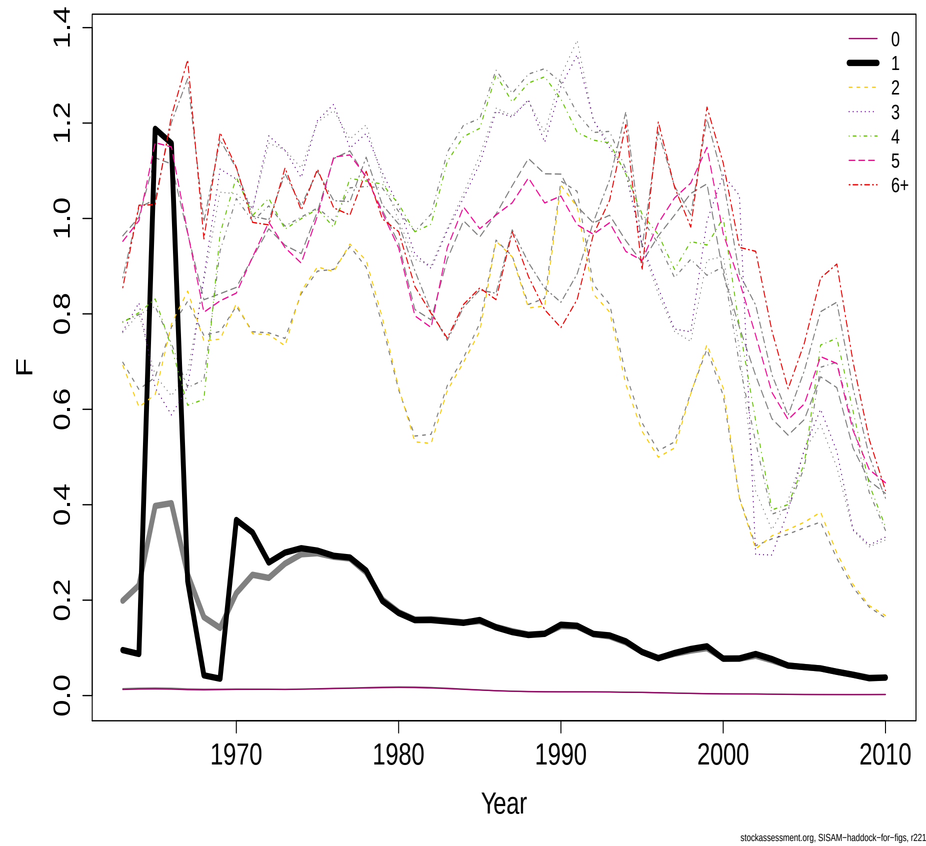
Robustifying w.r.t. recruitment spikes (Haddock)



- Comparing Gaussian (gray) with robust - no visual difference.
- Gaussian process assumptions were not restricting recruitment.

Robustifying w.r.t. fishing mortality (Haddock)





- Implies a big change in one years recruitment
- To accommodate the change in R, $F_{a=1}$ changed a lot in those years

So we could talk about

- Smoothing objectively or via ad-hoc settings.
- Robustifying

Northeast Atlantic Spurdog

Modified Punt-Walker: José De Oliveira, Jim Ellis, Helen Dobby

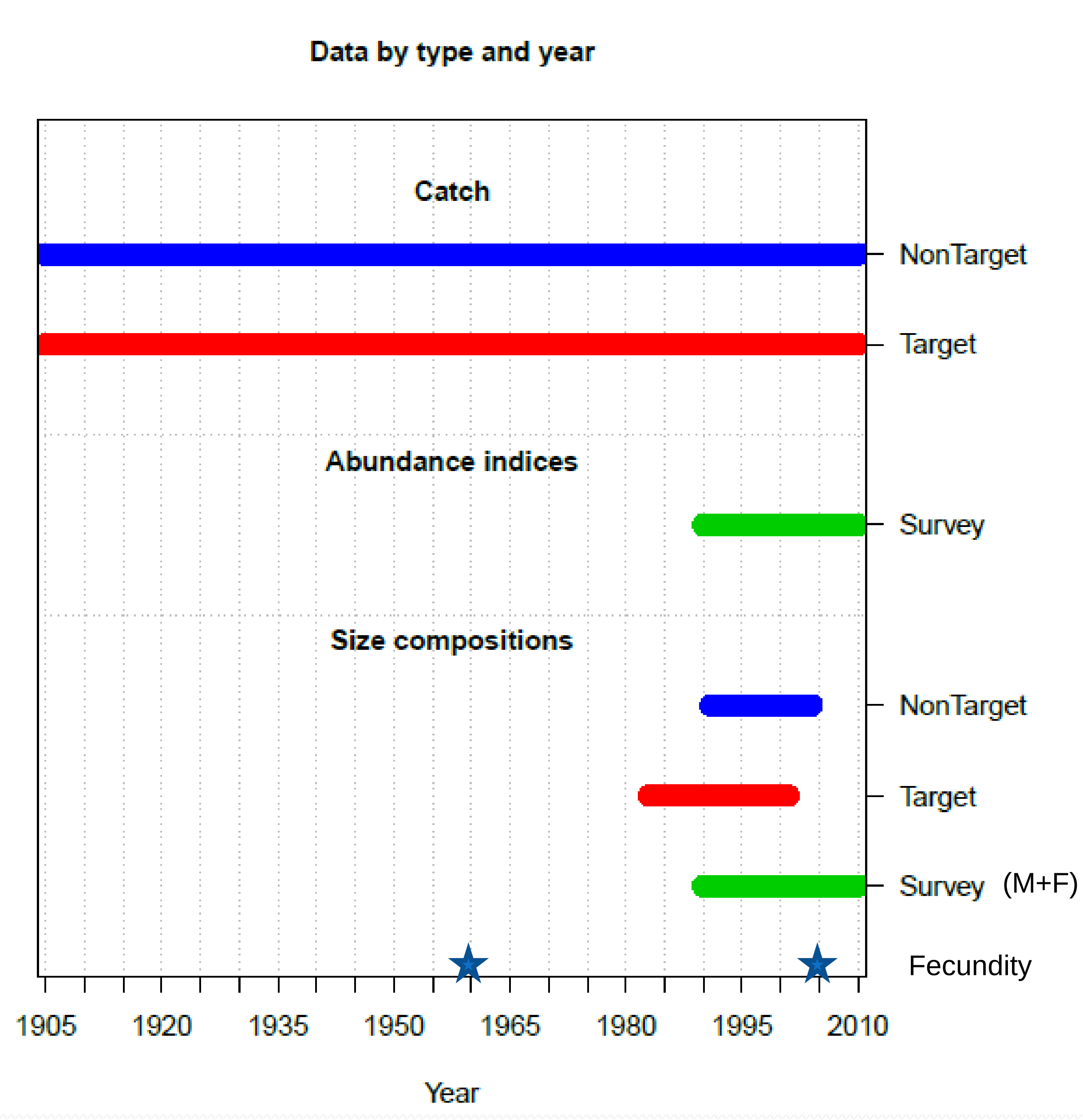
Modified AMAK: Pete Hulson

Stock Synthesis: Juan Valero, Rick Methot

Data

- Landings (1905-2010)
- Scottish commercial proportions-by-length category (1991-2004)
- English & Welsh commercial proportions-by-length category (1983-2001)
- Scottish survey proportions-by-length category (1990-2010) for males & females
- Scottish survey GLM-standardised CPUE and associated CVs (1990-2010)
- Fecundity data (1960 & 2005)

Data



Data by type and year
Catch
NonTarget
Target
Abundance indices
Survey
Size compositions
NonTarget
Target
Survey (M+F)
Fecundity
1905
1920
1935
1950
1965
1980
1995
2010
Year

Baseline assessment

- Based on Punt & Walker (1998) tope model
- Age- and sex-structured, but explicitly modelling length-based processes (maturity, pup production, growth, gear selectivity)
- Length-age relationship defines conversion from length to age
- Selectivity parameters estimated from proportion-by-length-category data

Pup production

- Pup production linked to number pregnant females, but annual deviations estimated

$$R_y = Q_y N_{pup,y} e^{\varepsilon_y}$$

$$Q_y = 1 + (Q_{fec} - 1)(1 - N_{pup,y} / R_0)$$

- Extent of density dependence in pup production Q_{fec} estimated using two periods of fecundity data (number pups per pregnant female): 1960 and 2005

Estimable parameters

- Pregnant females in virgin population (1)
- Survey selectivity (4)
- Commercial selectivity (2 x 3)
- Extent of density dependence Q_{fec} (1)
- Constrainted recruitment deviations 1960-2009 (50)

- Two fecundity parameters fixed based on scan of likelihood surface

Assumptions

- Model taken back to 1905 using landings data & assumptions about selectivity
  - Better reflect “virgin” conditions
  - Allow 1960 fecundity data to be fitted
  - Allow Q_{fec} to be estimated
- Two commercial fleets assumed, one with Scottish selectivity, and one with English & Welsh selectivity
  - Lack of data for other fleets
- Ignores discards (assumes 100% survival)

Length categories

16-31 cm (pups)

32-54 cm (juveniles)

55-69 cm (sub-adults)

70-84 cm (maturing fish)

85+ cm (mature fish)

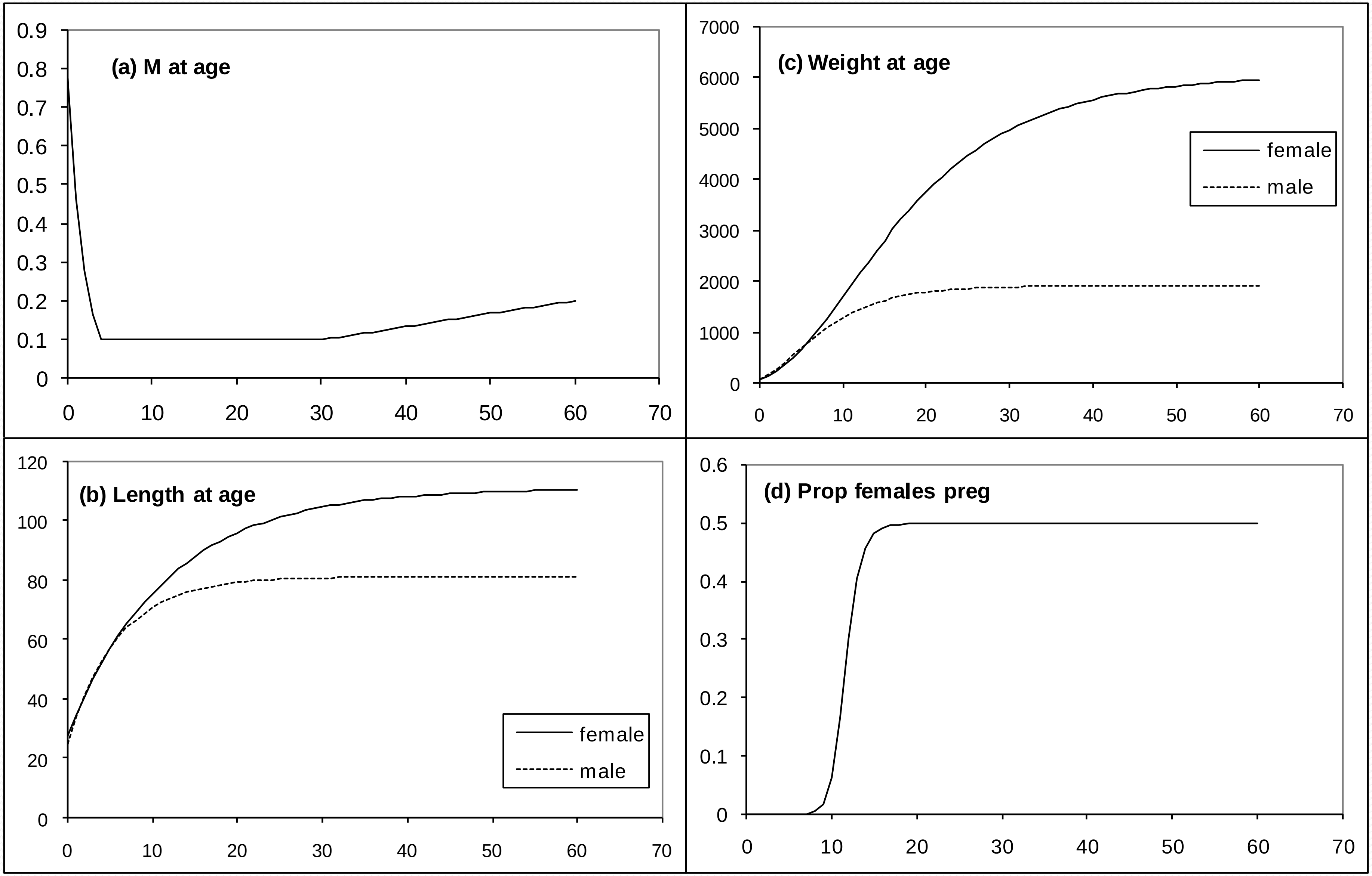
- Surveys:

  As above, but combine maturing and mature fish to form 70+ cm category

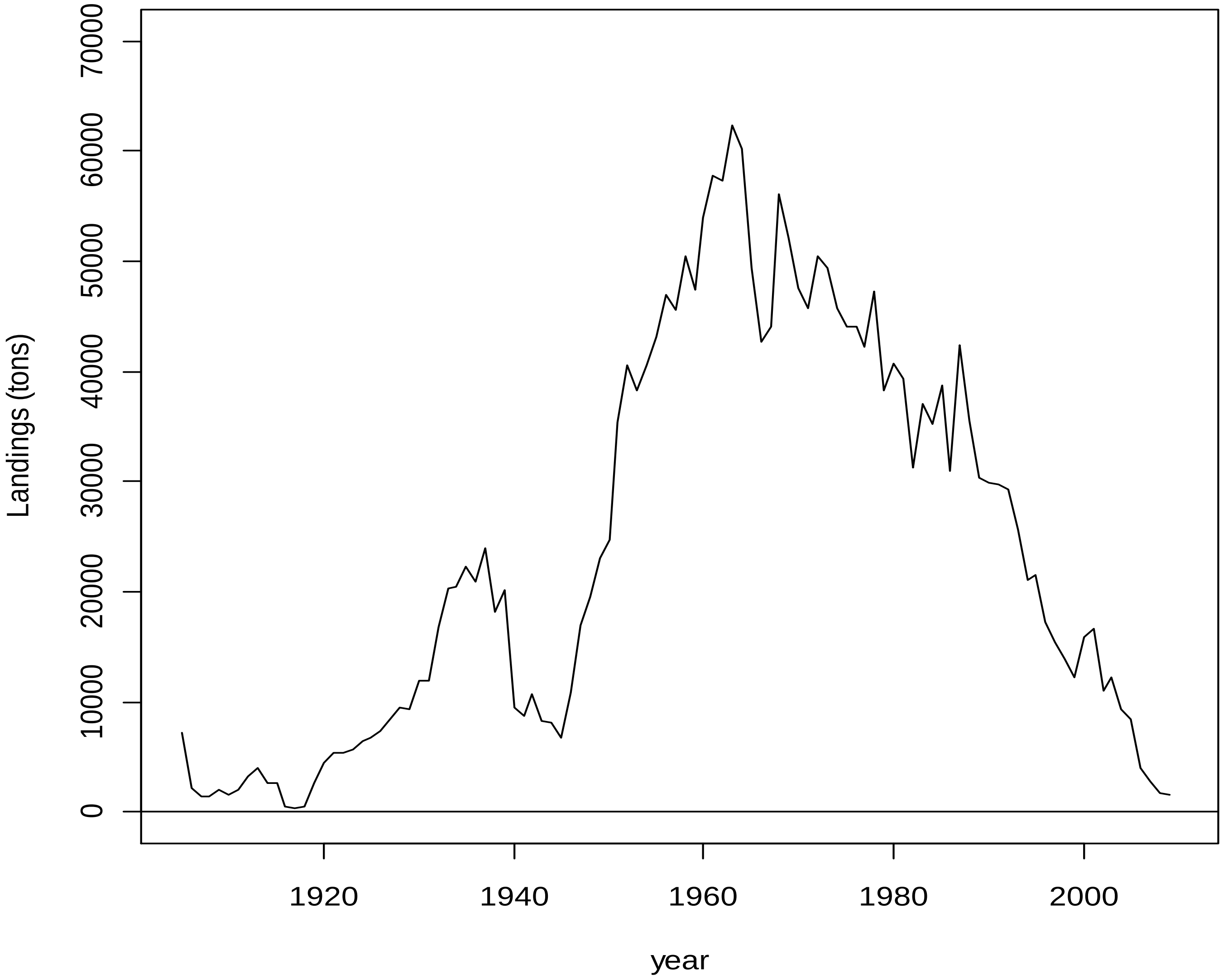
- Commercial landings:

  As above, but combine pups and juveniles to form 16-54 cm category

Life-history parameters



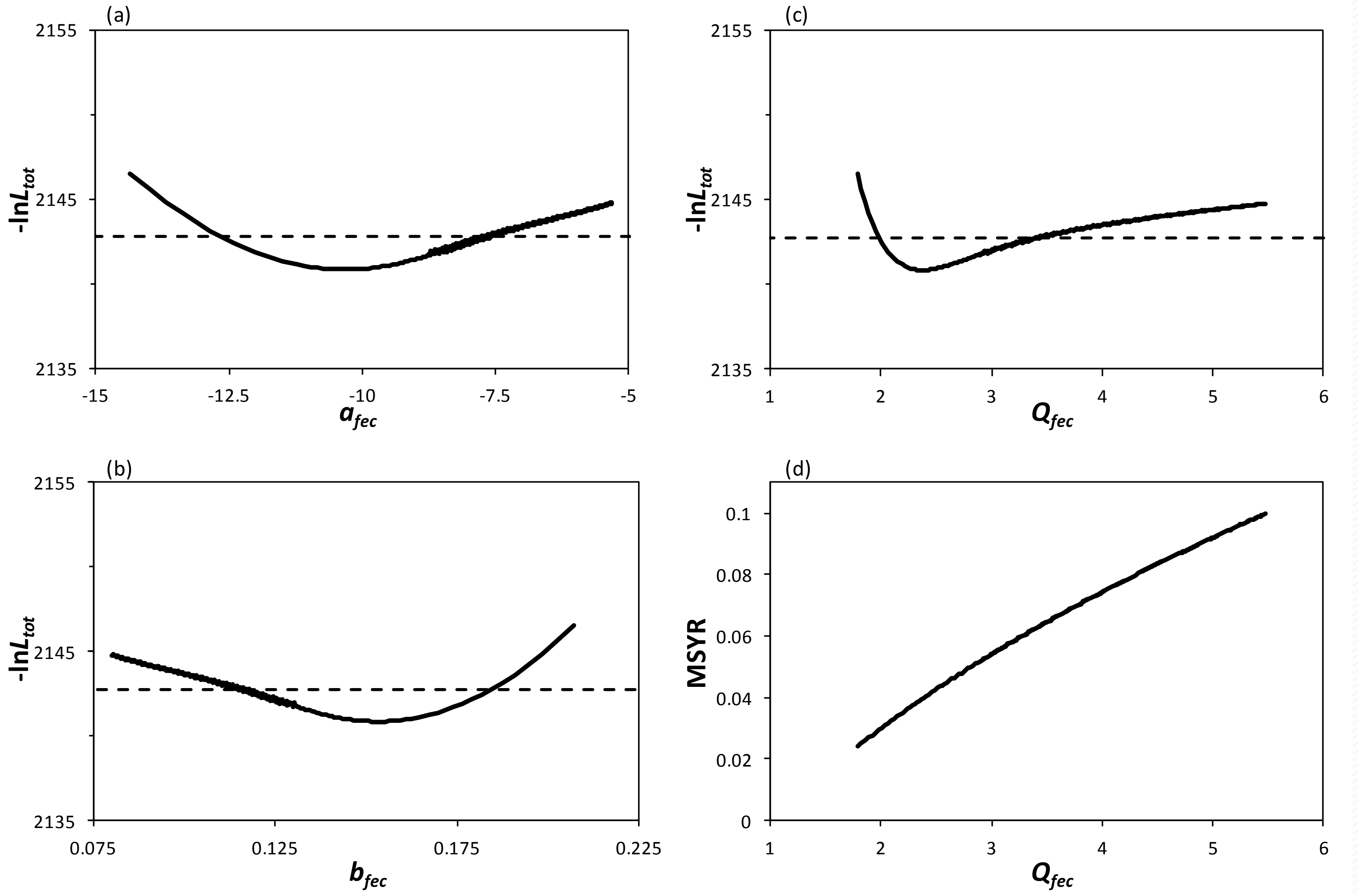
Landings



Landings (tons)
0
10000
20000
30000
40000
50000
60000
70000
1920
1940
1960
1980
2000
year

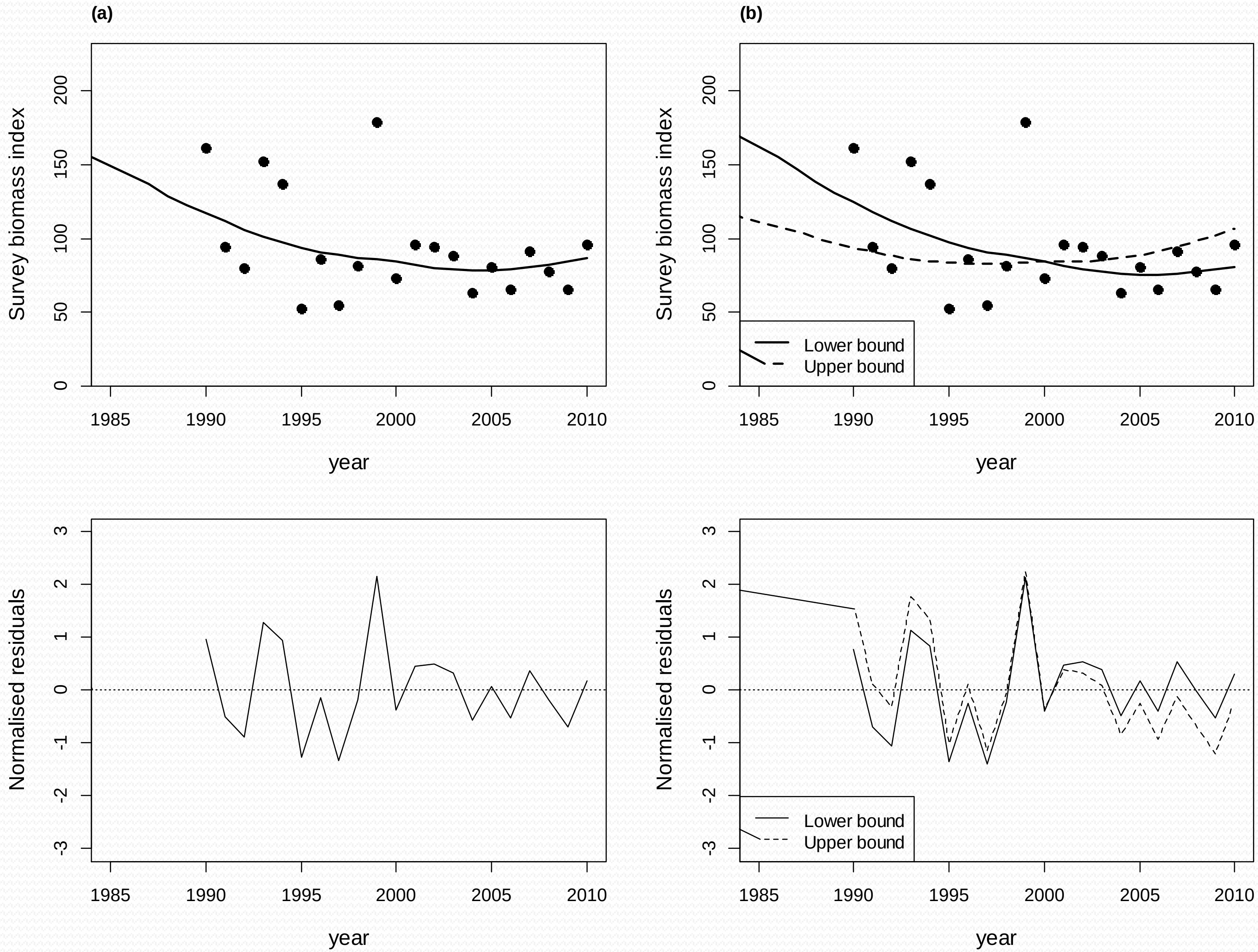
Model Fits

negative log-likelihood



Model Fits

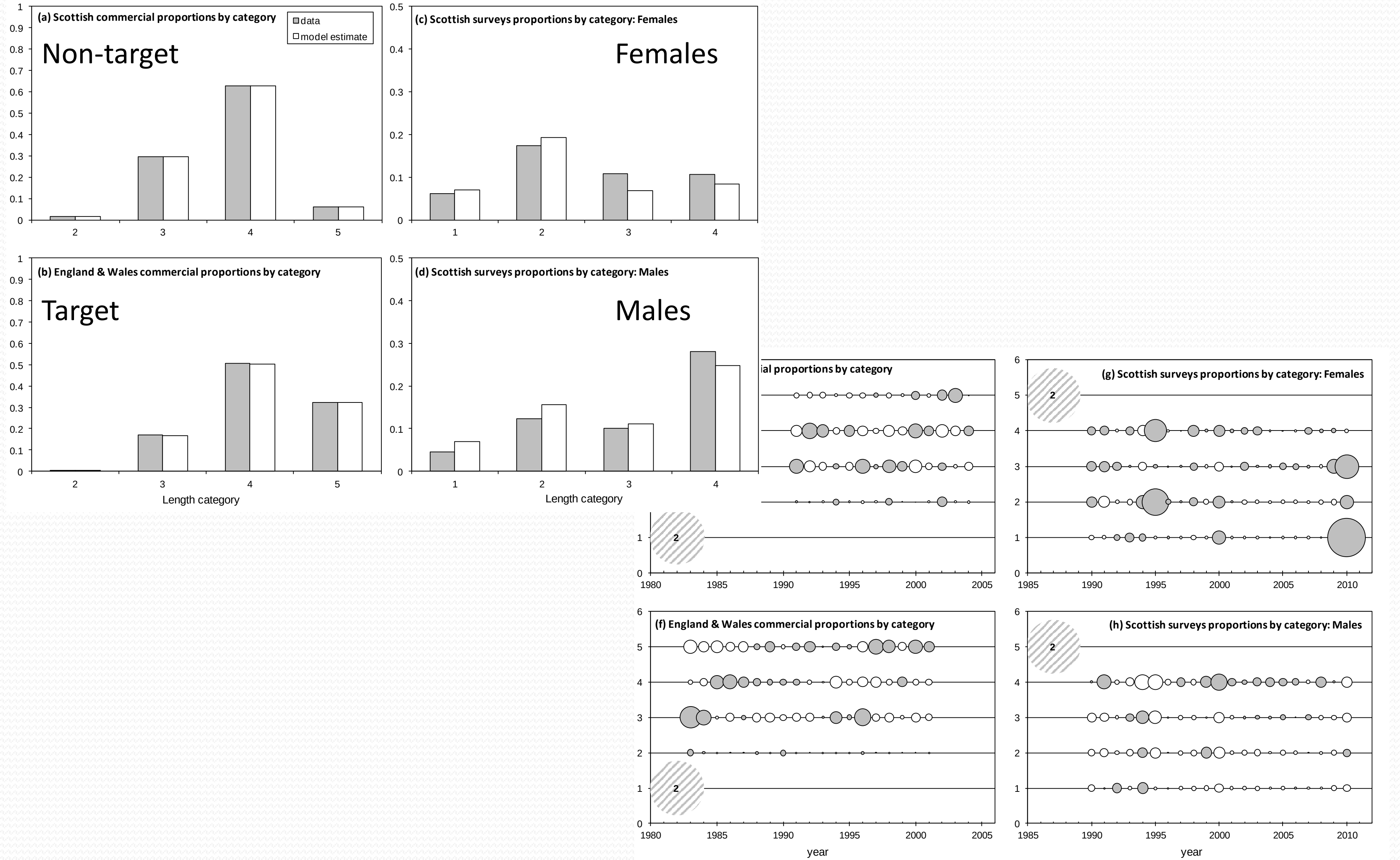
Scottish Survey Index



Model Fits Proportions by length

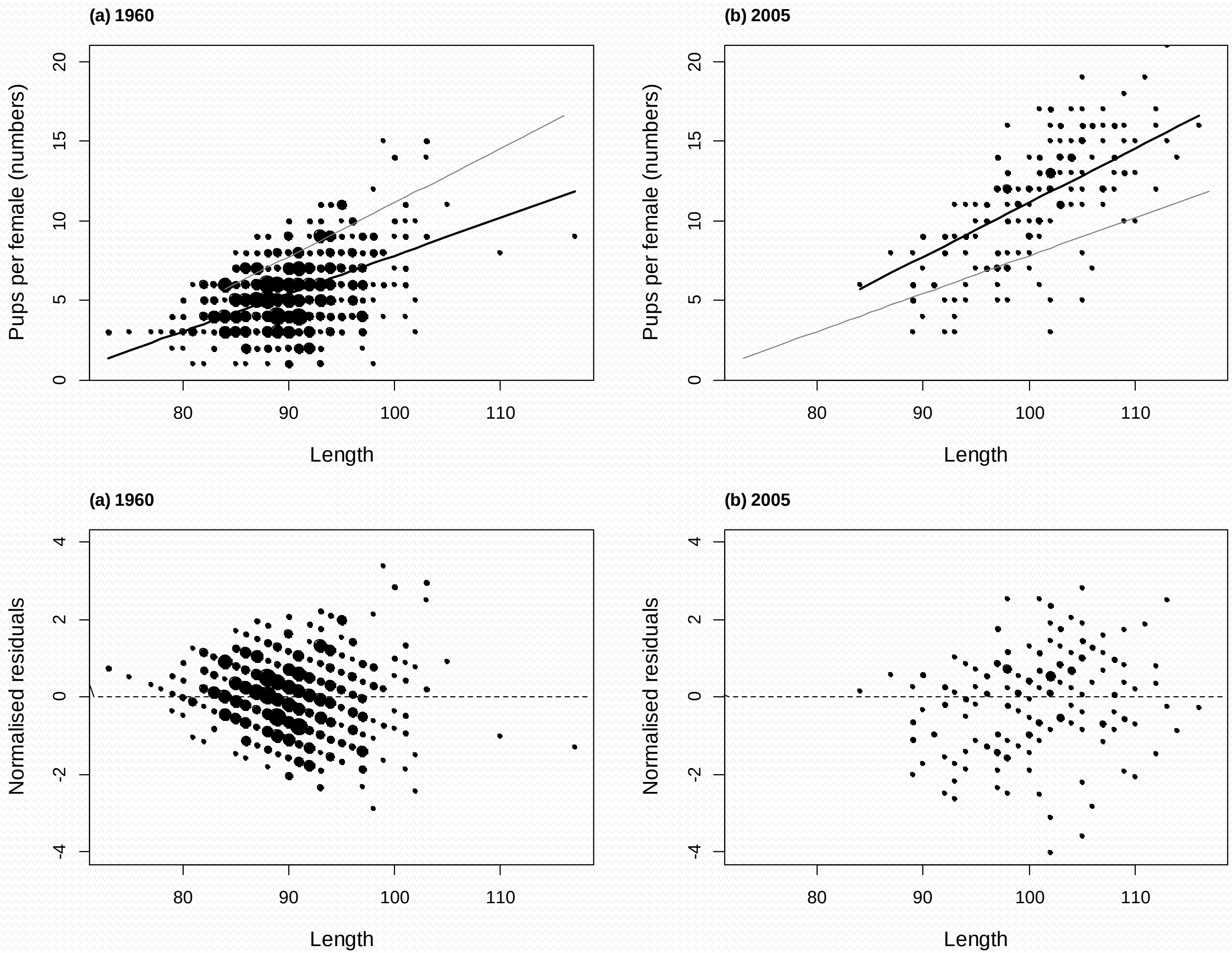
Commercial Catch

Scottish Survey



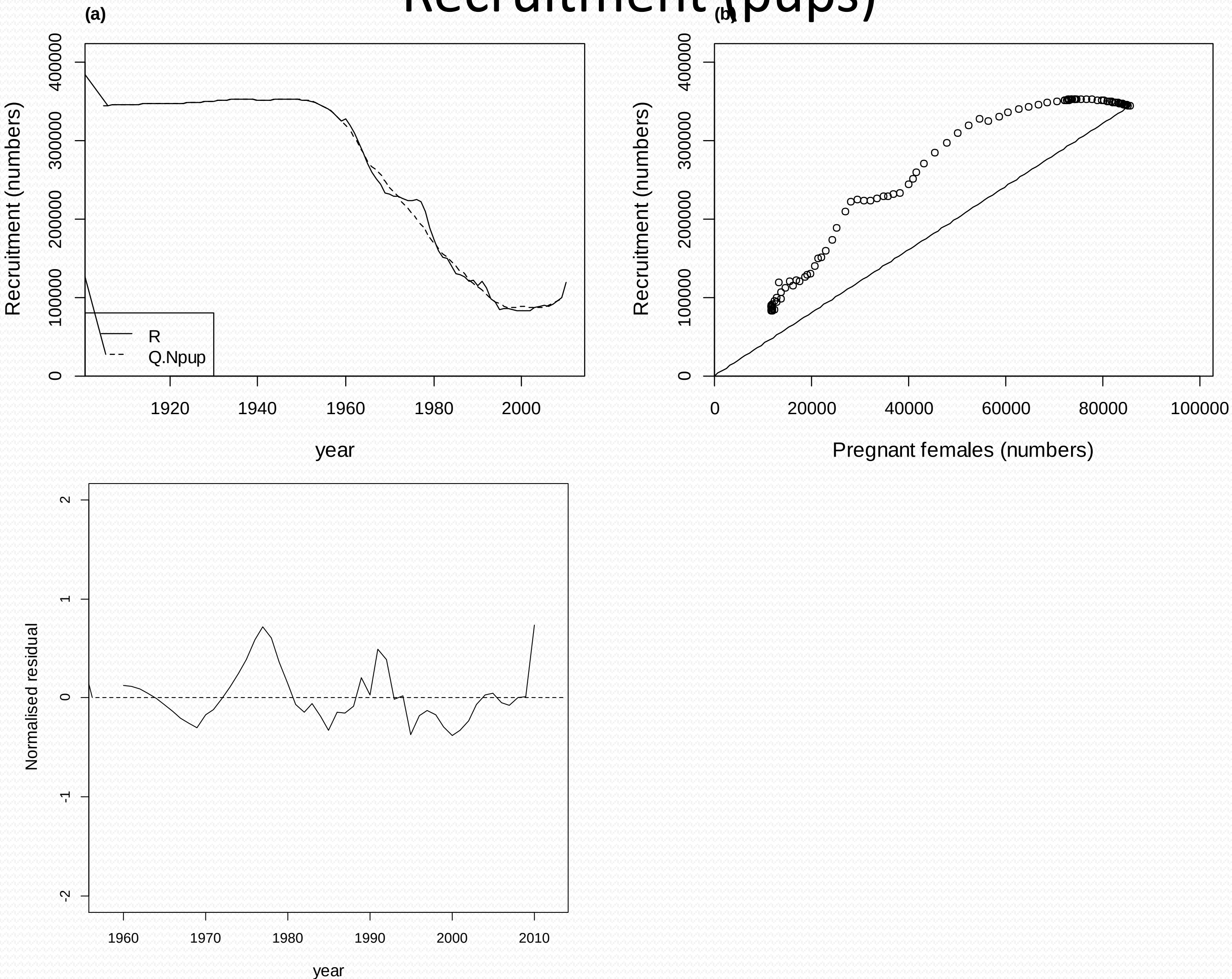
Model Fits

Fecundity data



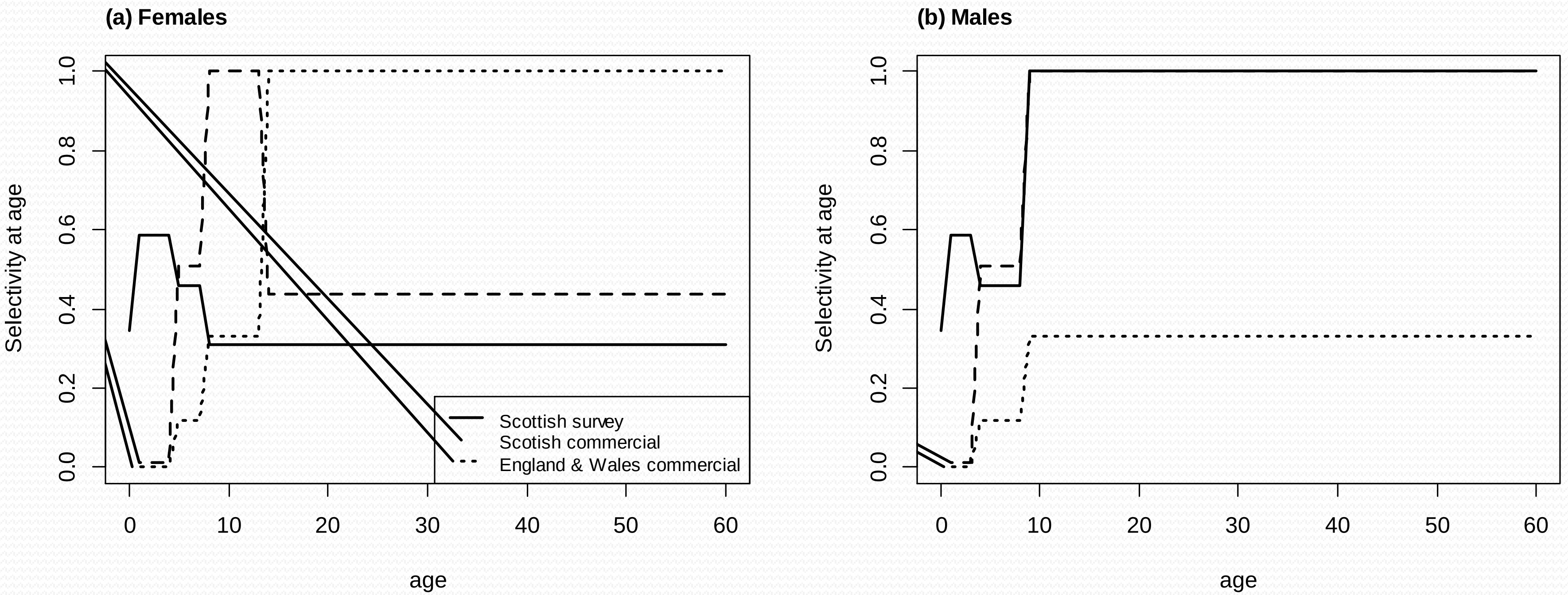
Model Fits

Recruitment (pups)



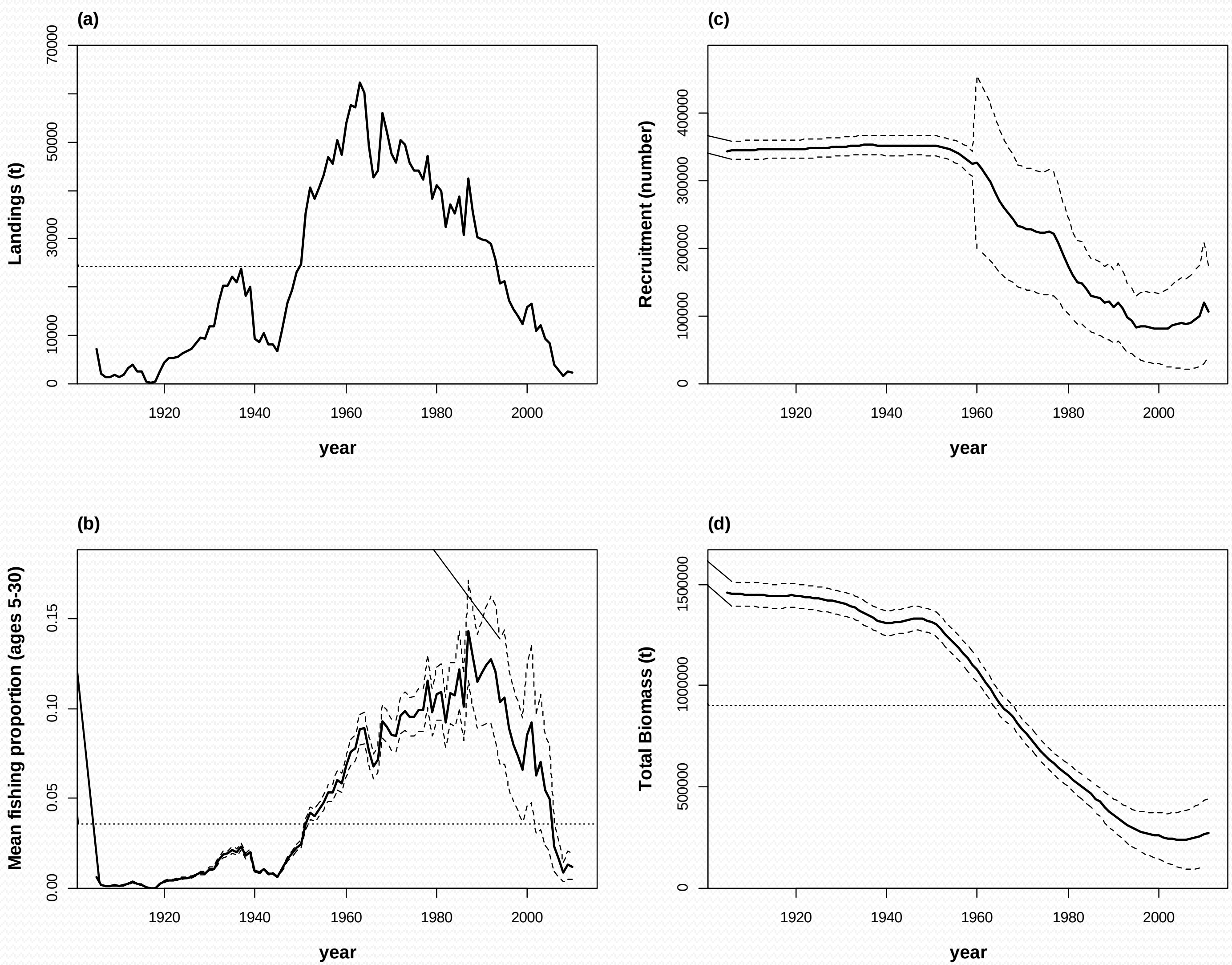
Model Estimates

Selectivity by length category

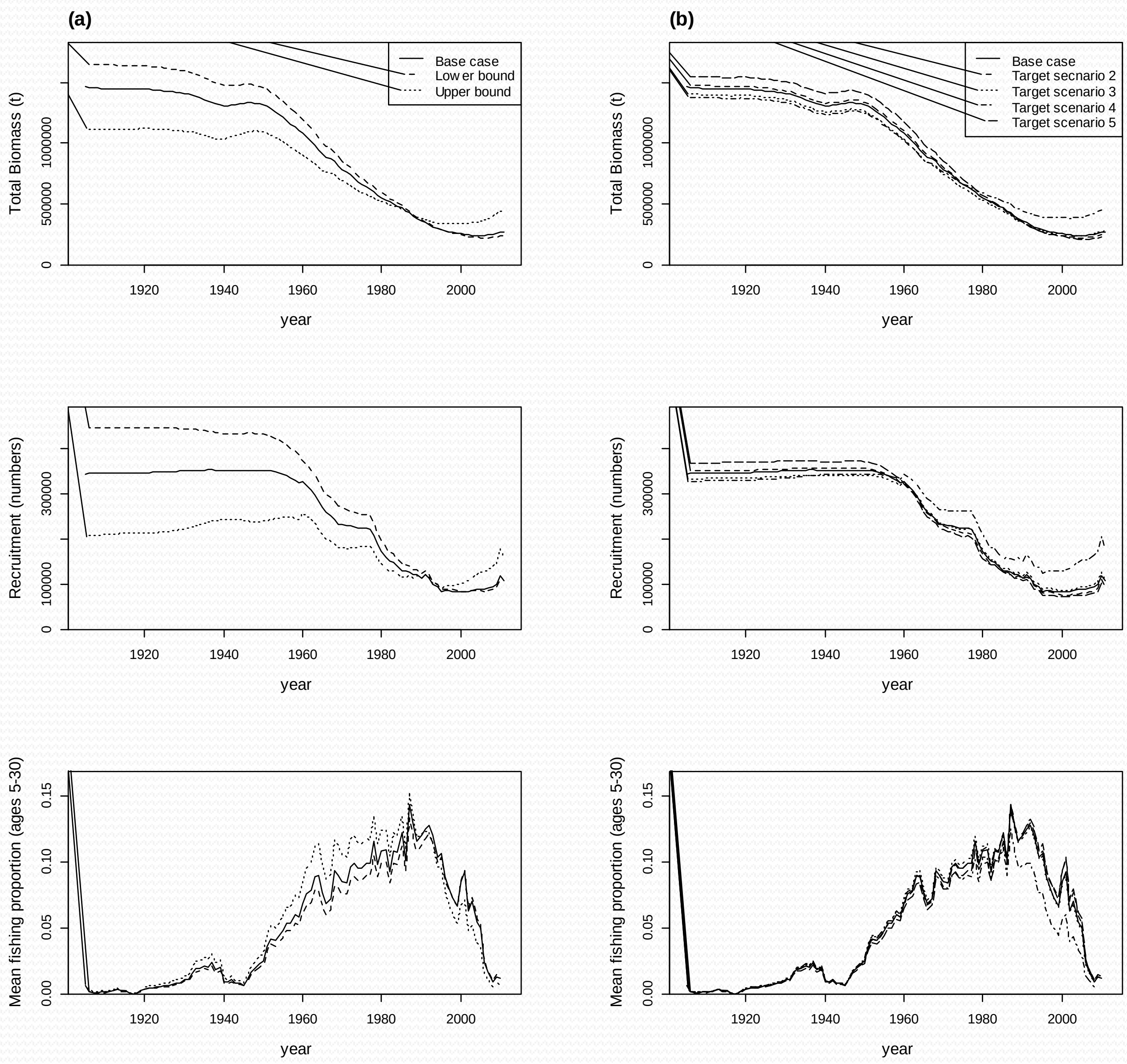


Model Estimates

Population trends



Model Sensitivity



Alternative Models

Modified AMAK
[Pete Hulson]

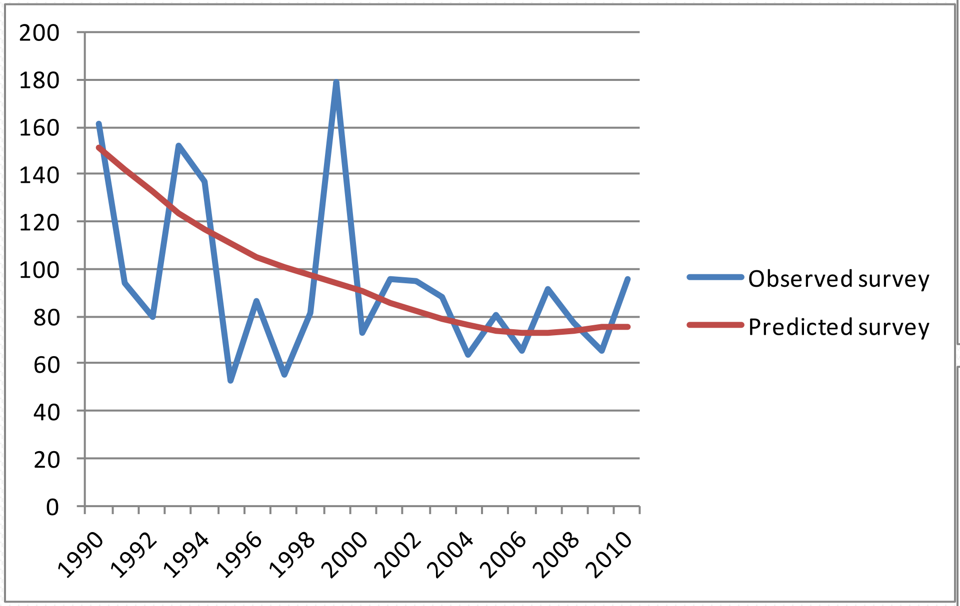
- Fecundity relationships not estimated
- Density dependence in pup production not considered
- Combined smallest length bins (16-31 & 32-54) for all data
- Modelled sex-specific target & non-target fishery length compositions, using proportions of males/females in survey
- Estimated annual F deviations

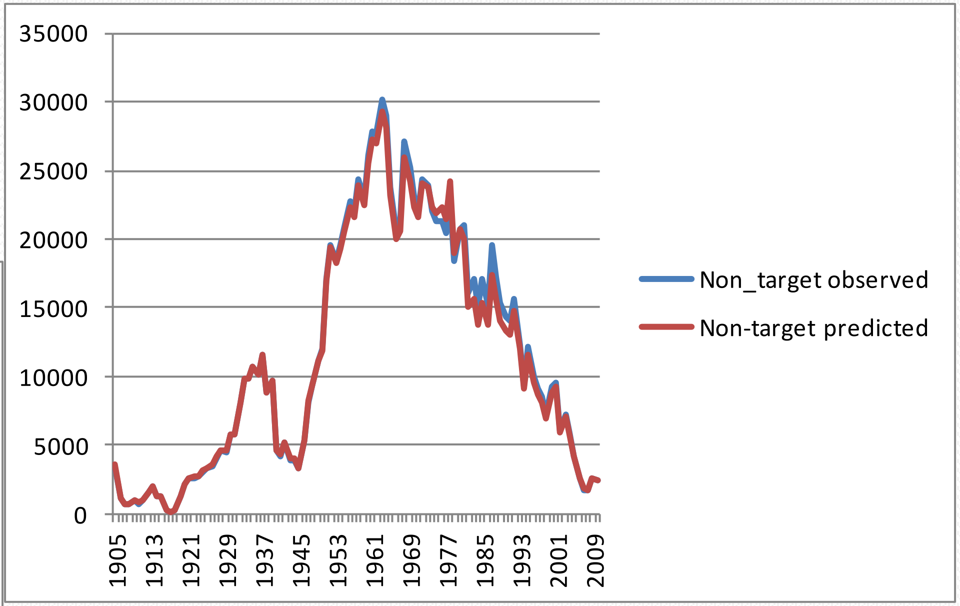
Stock Synthesis
[Juan Valero / Rick Methot]

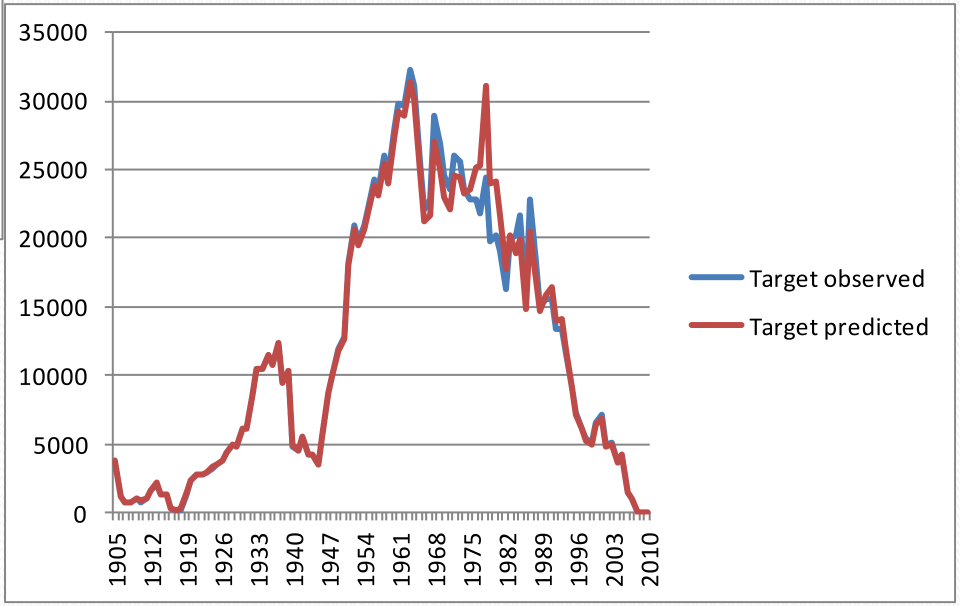
As close to original as possible

- No recruitment deviations
- Age-specific selectivity (14 parameters)

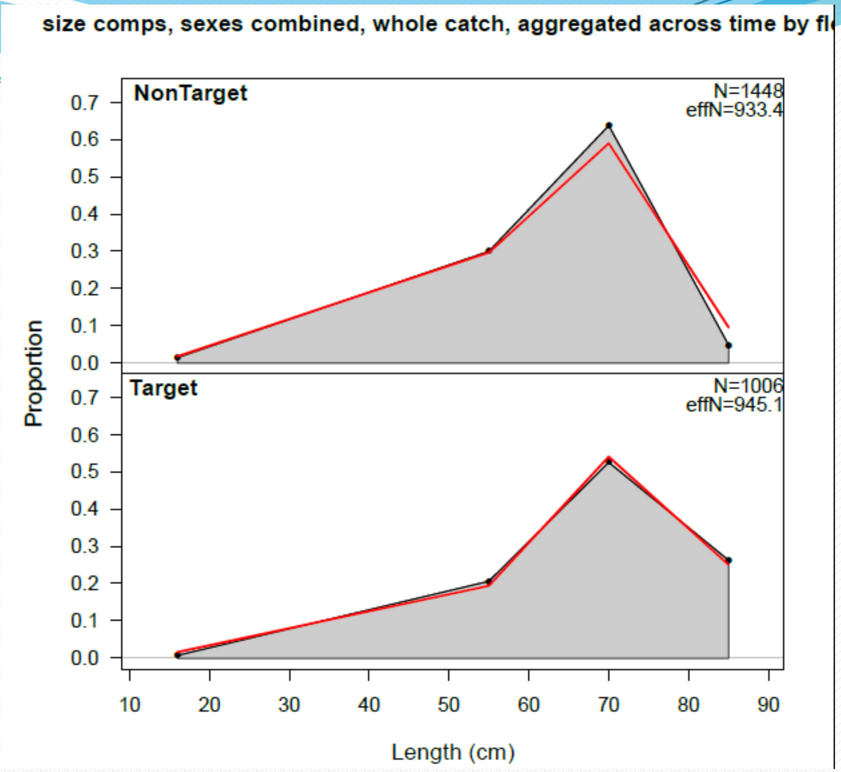
Modified AMAK

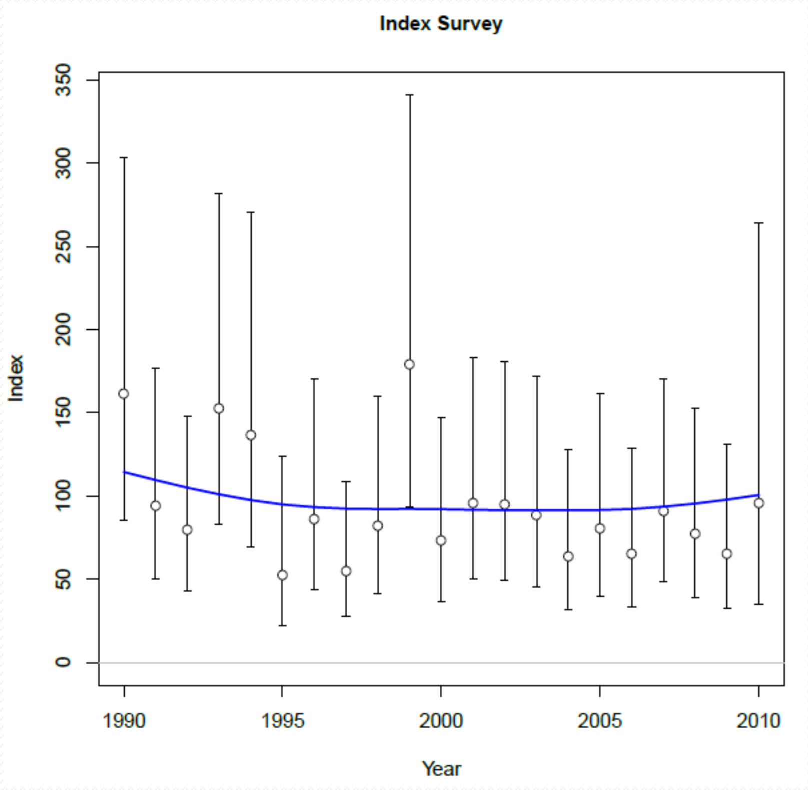


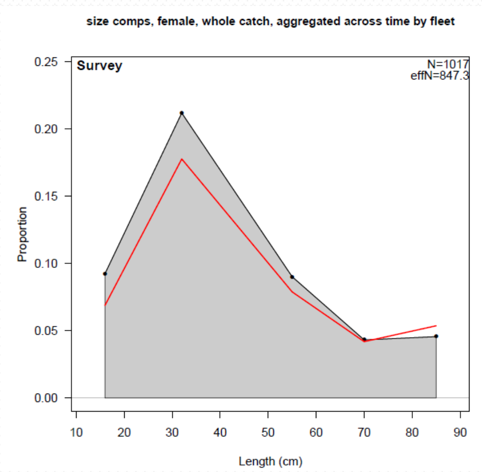


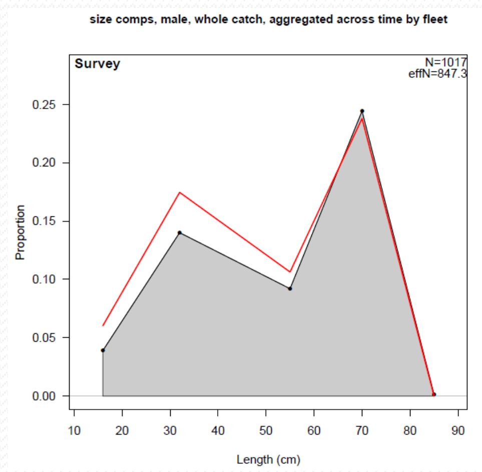


Stock Synthesis

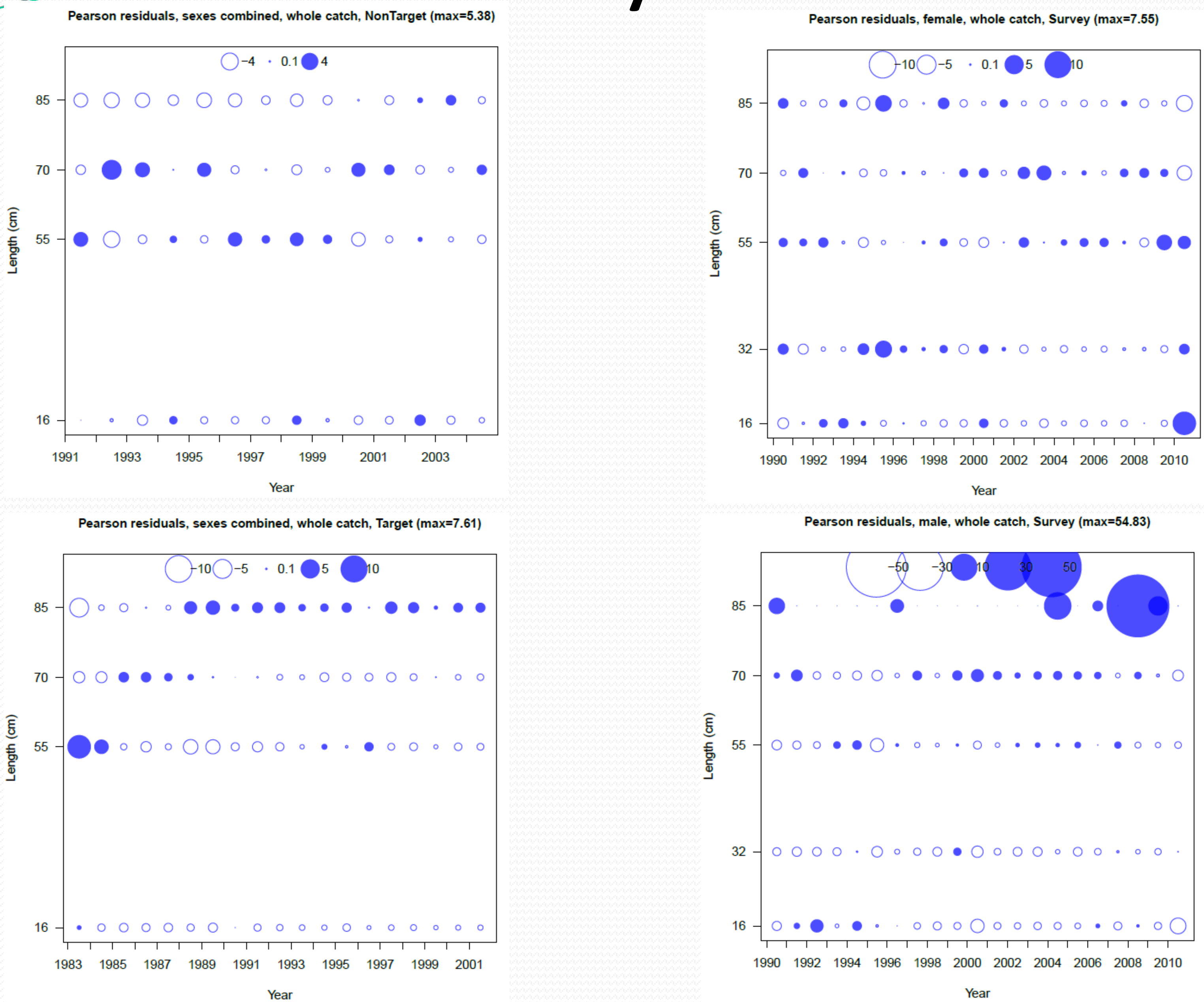




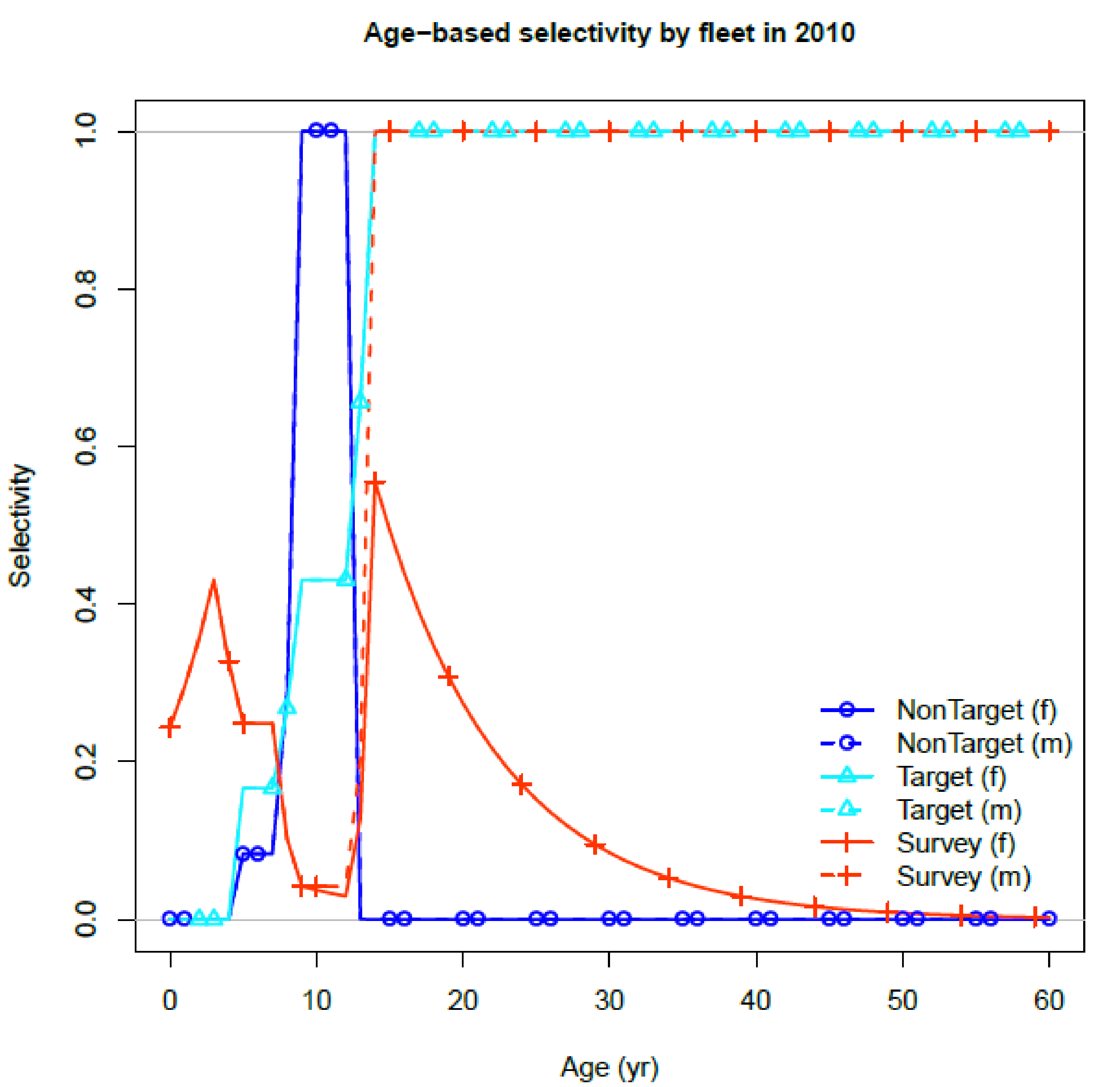




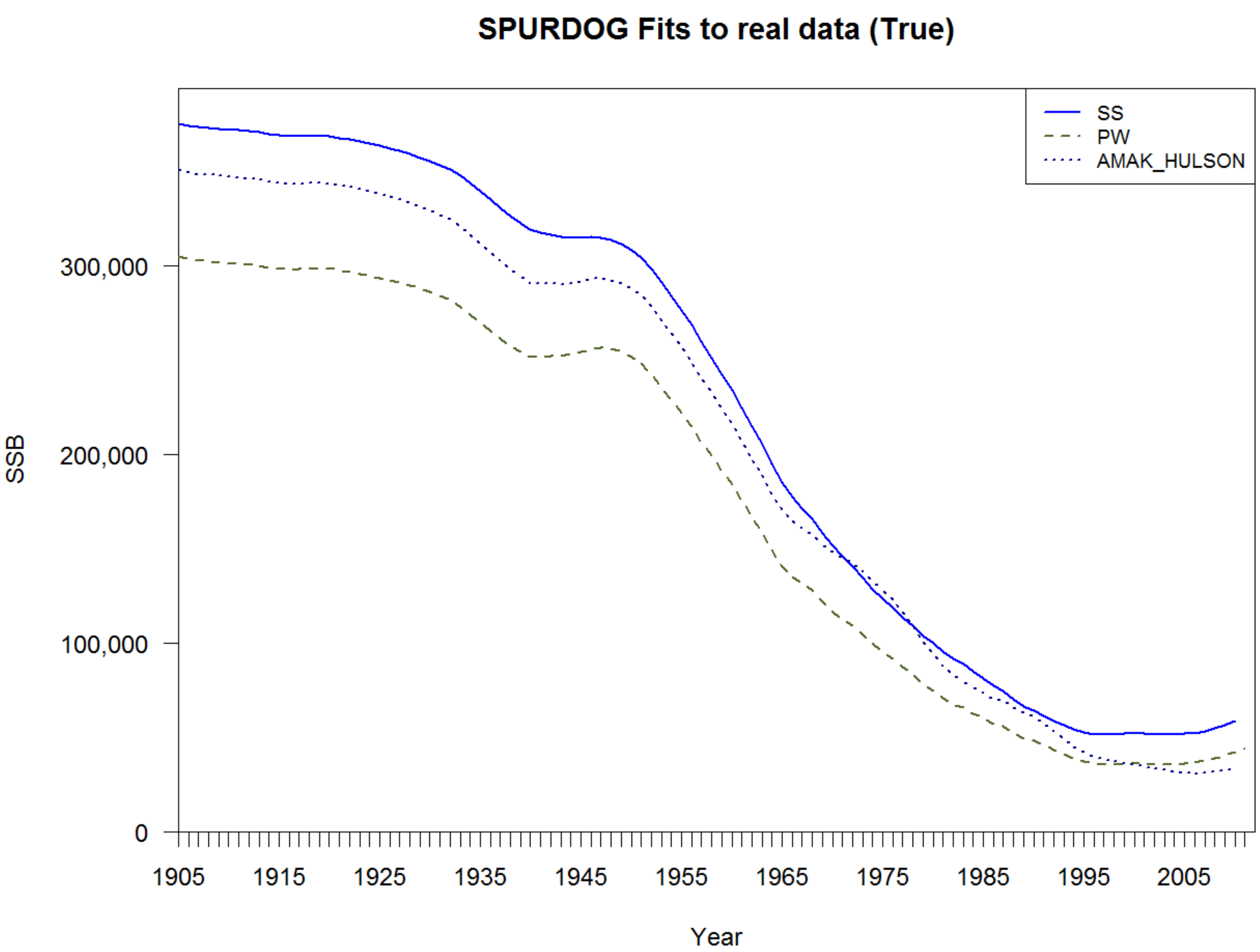
Stock Synthesis



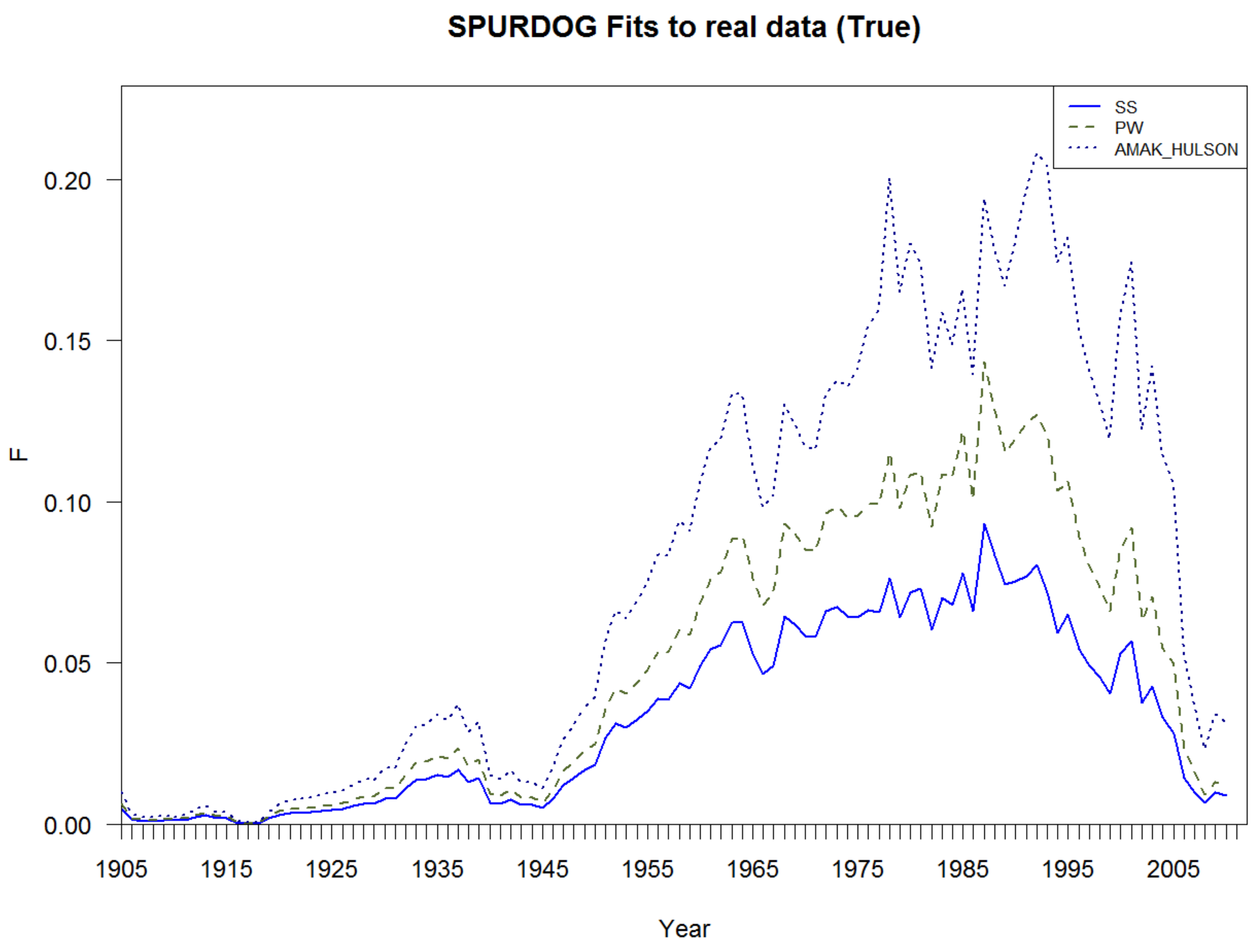
Stock Synthesis



Comparing models: SSB



Comparing models: F



Selectivity: theory, estimation, and application in fishery stock assessment models

Workshop Overview

P. R. Crone, M. N. Maunder, B. X. Semmens, J. L. Valero, and J. D. McDaniel

Center for the Advancement of
Population Assessment Methodology (CAPAM)
NOAA/IATTC/SIO
8901 La Jolla Shores Drive
La Jolla, CA 92037, USA

Major findings/high priority research areas

- Contact selectivity and availability
- General selectivity specification and estimation
- Asymptotic or dome-shape selectivity
- Size- or age-based selectivity
- Fleets as proxies for spatial processes
- Constant or time-varying selectivity
- Poor composition data
- Management strategy evaluations
- Standardizing selectivity in concert with CPUE estimation
- Survey selectivity
- Model selection and diagnostics

CAPAM – Selectivity Workshop

- Contact selectivity and availability = vulnerability
  - Vulnerability implicitly modeled in VPA and explicitly in integrated SCAA assessment approaches, but as a combined effect (single process)
  - Underlying processes/gear experiments provide some insight on expected shape of selectivity curve, but spatial processes (fish and fishery) can confound
  - ‘Outside the model’ estimates from empirical studies on gear operations are useful, but how to use information ‘inside the model’ not straightforward

CAPAM – Selectivity Workshop

- General selectivity specification and estimation
  - Disaggregate fishery composition data to aid investigating fleet and area variations in selectivity
  - ‘Unique’ shapes more common than thought, however impractical estimated curves warrant further attention
  - Ideally, selectivity approach should be flexible, robust (first, precise second), widely applicable, and allow for straightforward parameterization (*splines*)
  - Splines promising, research needed on number, location of knots, and performance
  - Don’t *a priori* rule out dome-shaped selectivity for any fishery
  - Time-varying selectivity seems more plausible assumption than constant
  - More uncertainty with length than age data and implications for choice of length- vs. age-based
  - Selectivity assumptions can lead to markedly different management advice

CAPAM – Selectivity Workshop

- Asymptotic or dome-shape selectivity
  - Debate continues
  - Some *domeness* expected, given simple mechanism of non-uniform fishing on spatially-segregated fish populations
    - Uncertainty surrounding selectivity of larger/older fish is most important for management
    - Selectivity of smaller/younger fish critical for accurately estimating relative cohort strength, and recruitment, and evaluating potential ecosystem changes (regime shifts)
  - Need for at least one asymptotic selectivity (fishery or survey) in assessment?
    - Common practice to stabilize parameter estimation
    - In principle, if only dome-shape selectivity, confounding with assumptions of *M*, and typically will increase uncertainty in abundance estimates
    - Asymptotic assumption likely more 'precautionary' (no cryptic biomass)—have objective stance for specifying at least one fishery/survey to be asymptotic

CAPAM – Selectivity Workshop

- Size- or age-based selectivity
  - Knowledge of mechanisms that influence both fishing process and fish biology
  - In age-structured assessment models, if age data used, either choice results in similar findings, however, choice is more influential if length data are relied on
  - Both could be operating simultaneously
  - If size-based, differences between mean length-at-age of catch vs. population, which will confound growth parameterization
  - Even if age data available, age-based selectivity may not be appropriate
    - e.g., large differences between adjacent age classes or over ages where size is generally similar (older fish)

CAPAM – Selectivity Workshop

- Fleets as proxies for spatial processes
    - Commonly used as proxy for spatial processes (fish and fishing)
    - Recent simulation-related research indicates that using fleet selectivity does not account for all bias in spatial structure
    - Ability for selectivity to explain potential spatial differences in age or size structure is likely fishery-specific

CAPAM – Selectivity Workshop

- Constant or time-varying selectivity
  - Quality and temporal coverage of fishery composition data are influential
  - Function of fishing and biological processes, unlikely to be homogeneous, temporally or spatially
  - Time-invariant selectivity likely masks real processes (fish and fishing) and can be overriding factor contributing to model misspecification
  - VPA
    - Requires complete and reliable age-composition data for all time periods, fleets, and surveys, which is often not possible for stocks of interest
    - Selectivity is derived quantity from calculated *F*-at-age within each year and often is highly variable from one year to next (implicit selectivity parameterization only)
    - Plus-group estimation is restrictive and results can be sensitive to assumption (e.g., stocks with low total mortality)
  - Periodicity of time-varying selectivity
    - First, implement/evaluate at finest scale (annual or random walk), then use extended time blocks accordingly
    - For potential time blocks, determine breakpoints objectively (as possible), e.g., notable changes in fishery regulations, residual analysis of fits to composition
    - Sensitivity analysis (experiment framework) for plausible alternative blocks
    - Evaluate likelihood profiles of a population scaling parameter (virgin recruitment, catchability) based on alternative blocks
  - Bottom-line: ignoring temporal changes in selectivity can produce biased estimates of management quantities and underestimate uncertainty

CAPAM – Selectivity Workshop

- Poor composition data
  - Most of input data and estimated error in typical integrated stock assessments associated with model fits to fishery composition data
  - Composition data that strongly influence assessment results may indicate misspecified selectivity
  - Effect of composition data on fit to population indices of abundance should be closely examined
    - ‘Down-weight’ composition data relative to abundance data—do not let size/age time series override CPUE/survey index
    - Evaluate residual patterns associated with fits to abundance indices, i.e., in-line with assumed variances?
  - Poor fits to composition data more likely caused by poor model assumptions than less than ideal data-weighting schemes
  - Do not have blind faith in (available) age data
    - Production ageing labs need to stay abreast of most robust, precise, efficient age-determination methods
    - Compare/contrast how length data perform in assessment model as well

- Management strategy evaluations
  - Formal analysis to evaluate catch-determination methods over a broad range of model scenarios (‘states of nature’)
  - Alternative selectivity assumptions/parameterizations could be easily accommodated in MSE framework
    - Alternative options tested during the estimation/catch specification stage using plausible scenarios from the operating model
    - Meaningful tests include age vs. length data, age- vs. length-based selectivity, parametric vs. non-parametric forms, dome-shape vs. asymptotic, constant vs. time-varying
    - MSE should provide confidence that selectivity specification is robust in terms of management outcomes and risk

CAPAM – Selectivity Workshop

- Survey selectivity
  - Needs special consideration, given ideally, should be designed in a manner consistent with assumption that selectivity is constant and asymptotic
    - Standardized survey gear/protocols may approximately achieve constant (contact) selectivity, but constant availability unlikely for many fish populations
    - Standardized, reliable, and representative composition data from surveys are informative for robust assessments, even in applications when fishery selectivity is assumed to be time-varying or composition data from the fisheries are down-weighted
    - If fishery/CPUE selectivity (and catchability) assumed to be changing over time, then would be remiss not to investigate appropriateness for surveys

CAPAM – Selectivity Workshop

- Model selection and diagnostics
  - Model selection and evaluation are not straightforward for any selectivity parameterization
  - Classical statistical tests for formal model selection may not be useful (valid) in most stock assessments, given multiple sources of data and correctly specifying likelihood specifications, sample sizes and variances, and random effects—do not over-interpret
  - Presently, using simulation analysis for many forensic investigations (including selectivity) surrounding assessments is gaining popularity and appears promising
  - Profile likelihoods involving scaling parameter (R_0)
  - Bottom-line: little formal guidance on good, better, best diagnostic tool to employ—group felt one of the highest priorities for future research

CAPAM – Selectivity Workshop

Future work

- Continue with selectivity research, including splines, composition/length vs. age, data weighting, VPA - spatial F/selectivity form
  - Establish working group / begin synthesis and documentation related to *Good Practices Guide*
  - Visiting scientist research
- Begin related research projects for *GPG* (e.g., modeling growth in stock assessments)
  - Prepare for growth workshop (late 2014)
- Conduct classes/short courses—SIO and international
- Build on momentum to link formally with institutions/programs involved in similar research—regionally, nationally, and internationally
  - Stock assessment modeling issues
  - ADMB Project
  - SS model development
  - Joint workshops (national and international)
  - *Next generation of stock assessment scientists*
- Conduct classes/short courses—SIO and international agencies/institutions
- Visit **www.CAPAMresearch.org**

NOAA FISHERIES

A Generalized Assessment Model to Obtain Consistent Management Advice from Diverse Data

Richard D. Methot Jr.

Science Advisor for Stock Assessments

World Conference on Stock Assessment Methods

Boston, MA

July 17, 2013

Stock Assessment Goals

- What harvest policy is sustainable and provides balance between preventing overfishing and attaining maximum fishing opportunities?
- Does current level of fishing (F) exceed that policy?
- Has abundance (B) been so reduced by past fishing as to put the stock and ecosystem at risk?
- What future catch would implement the policy?

Assessment Data and Situations

DATA

- Catch only
- Catch and stock abundance
- Catch, abundance and/or composition
- Add ecosystem/ climate/ habitat factors

SITUATIONS

- Short time series vs. long-term series containing contrast
- High F vs. low F
- Stable biology vs. environ/eco driven changes in process
- Degree of stock fluctuations (M + sigmaR)
- Degree of spatial viscosity

Assessment Approaches

- Catch Only
  - Time series, no biology
- Biomass Dynamics
  - Simple tuning factor
  - Time series tuning
  - STATISTICS: measurement vs. process error
- Age and/or Size Structured
  - Noisy data with gaps
  - Full catch-at-age
  - STATISTICS: Penalized pseudo-likelihood, Integration across random effects, Kalman filter
- Multi-Species with M and/or technological linkages

Added Features:
Spatial
Multi-species
Covariates

Biomass vs. Age Model Dichotomy

Biomass Dynamics
r, K parameters

Age-Structured
Empirical Reconstruction;
Then Spawn-Recruit

F_{msy} gives B_{msy}
near 0.5*K

$\neq$

F_{msy} gives B_{msy} near 0.3*K,
or lower.

- Use 3-parameter forms that align these approaches;
- Don't ignore effects on SSB when using F_{max} as F_{msy} proxy

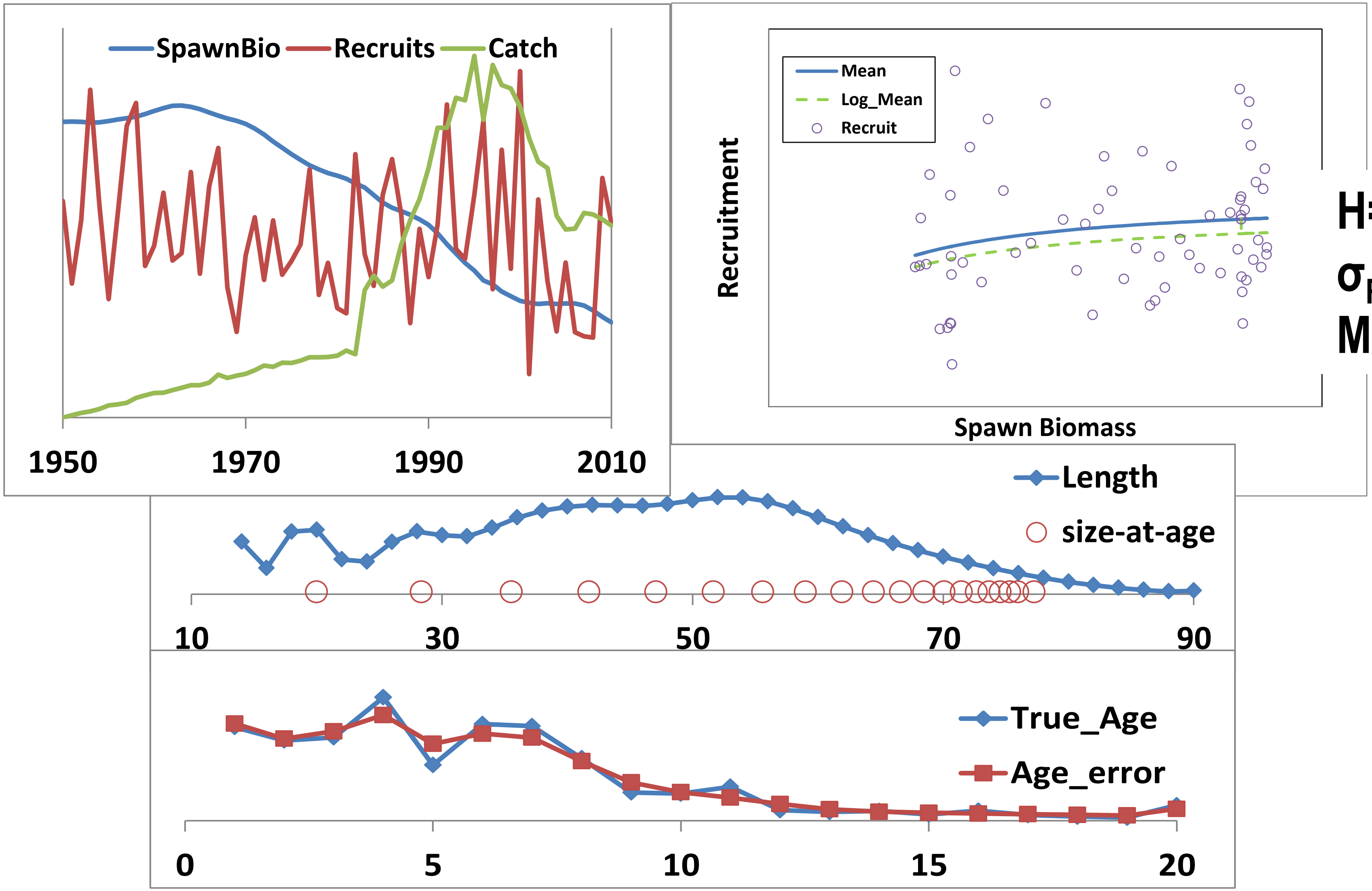
Desirable Model Characteristics

- Measures F, B, and productivity
- Estimates reference points and does forecasting
- Assimilates diverse types of data
- Consistency (no dichotomy as on previous slide)
- Statistically rigorous
- Biologically realistic
- Responsive to time-varying ecosystem/environmental processes
- Easy to use; includes A.I. to guide good usage practices
- Spatial
- Multi-species

How do Data Influence Assessment Results in a Generalized Model – Stock Synthesis?

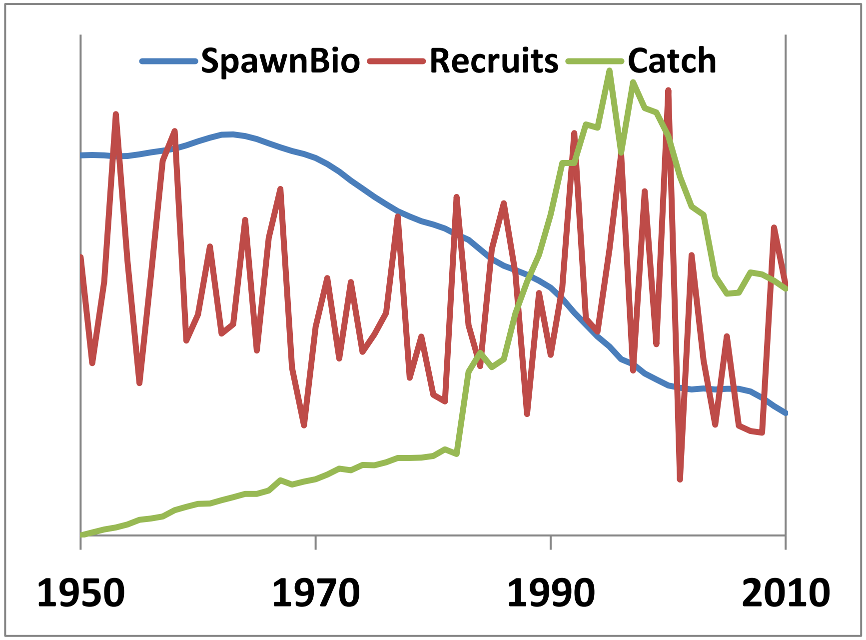
- Consider three data situation
  1. Scalar observation at end of time series
     - Mean length
     - Current F
     - $B_{current} / B_0$
  2. Time series of relative abundance
  3. Composition data
     - Perfectly precise ages
     - Ages with ageing imprecision
     - Lengths

Example Simulated Population



H=0.7;
σ_R=0.4;
M=0.2

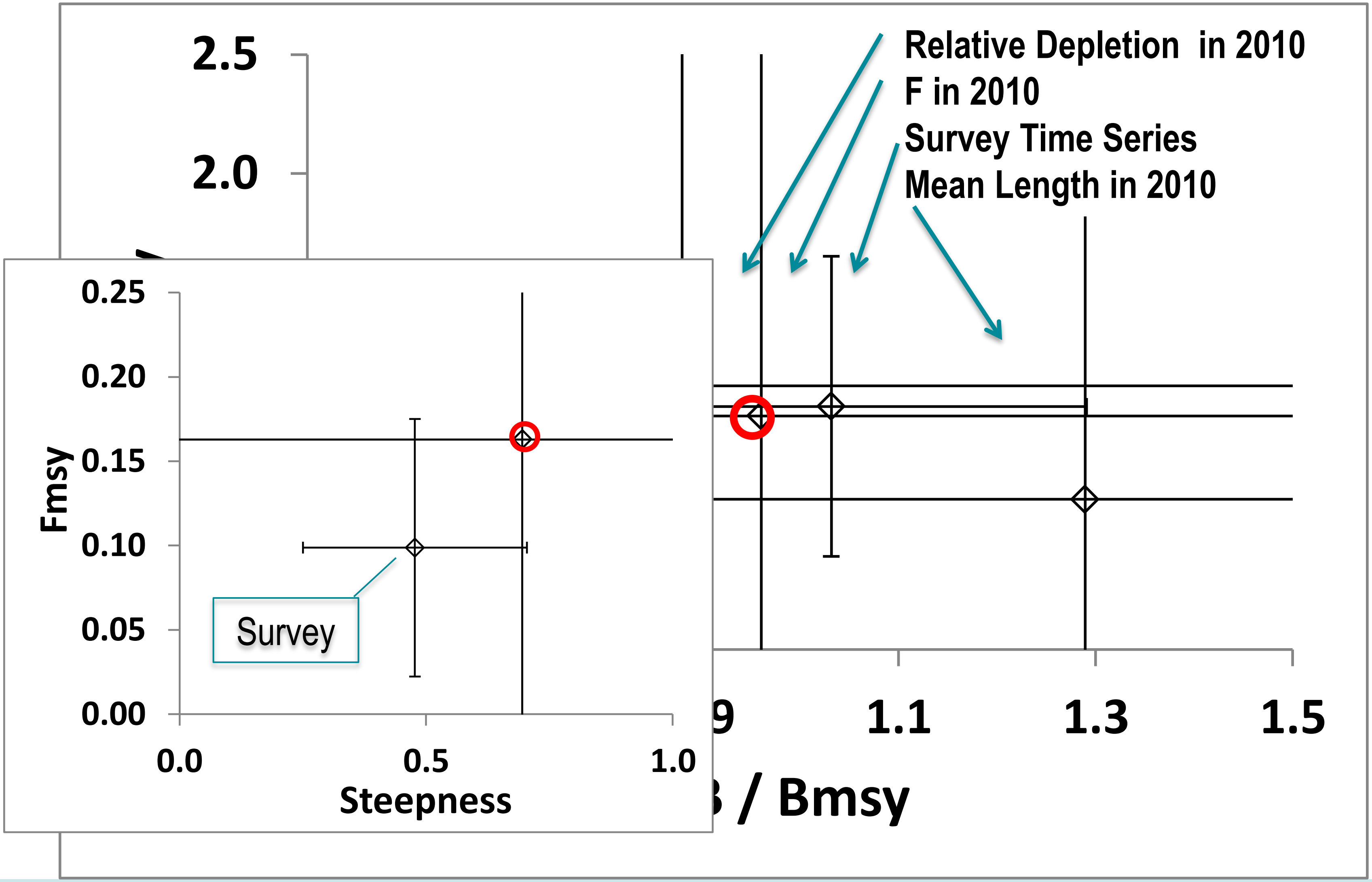
Generate and Analyze Simulated Data Using Stock Synthesis



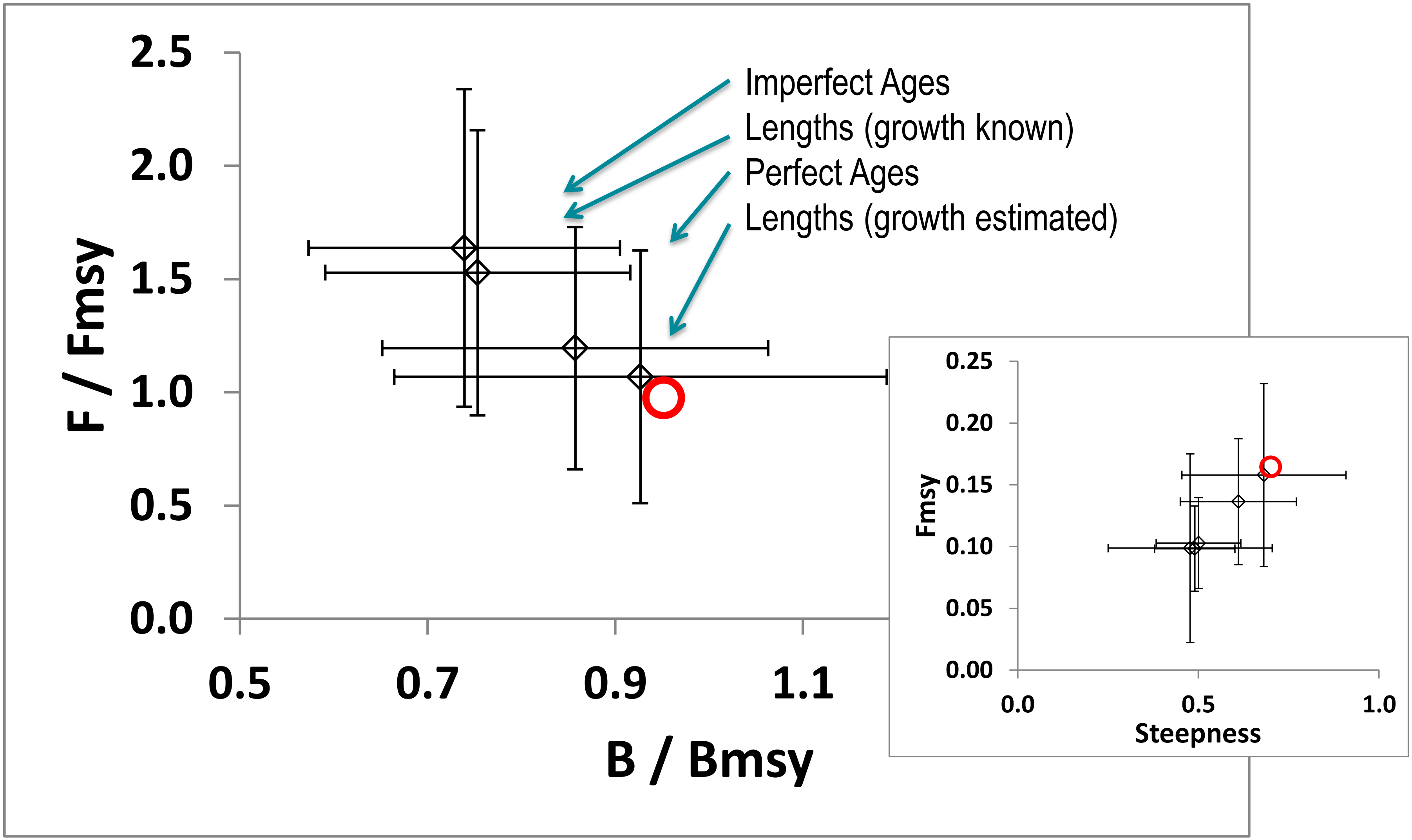
- Fishery age, length, and imperfect ages beginning in 1971
- Survey of spawning biomass beginning in 1981
- Various scalar measures in 2010

- Analyze each data scenario using Stock Synthesis (SS)
- Allow estimation of some or all of:
  - Steepness
  - Selectivity
  - M
  - Recruitment deviations
  - Growth
- Use informative priors in a penalized likelihood framework
- Focus on variance of model results

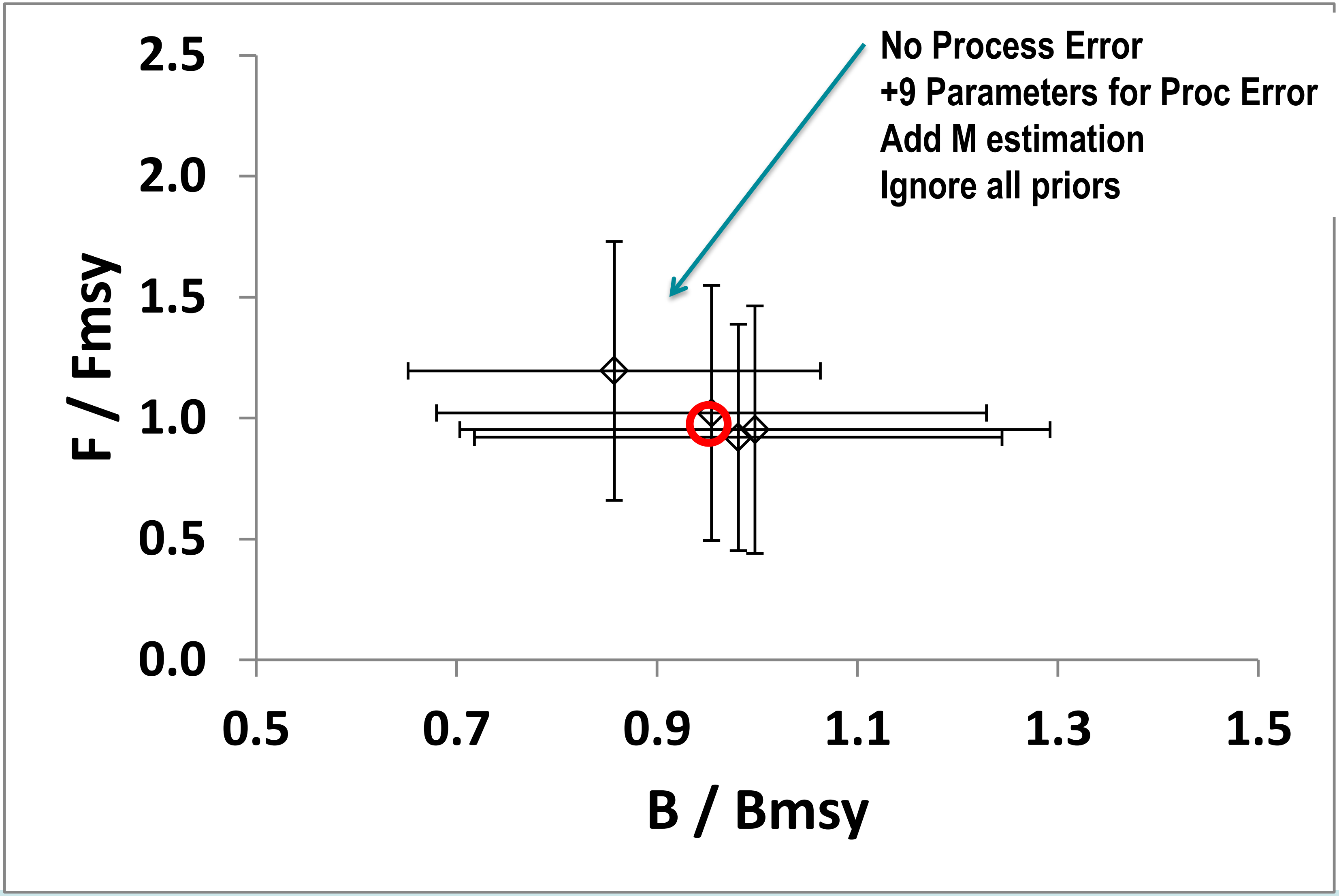
Results with Simple Data



Results with Composition and Survey Data



Age Data and Survey; Est. Selec. Process Error



Simulation Summary

- Catch time series plus some simple indicator of F is highly informative
- Three types of composition data ~ equally informative
  - Truly random data
  - Repeated observations of each cohort
- Adding process error in estimation did not greatly degrade precision

- **A generalized model enables blending information from diverse data and making comparisons such as this**
- **Lightly informative priors are important part of approach**
- **Real data must be much worse than random measurement error**

Other Simulation Studies

- Fidelity of M and h estimation in assessment models (Lee, Piner, Maunder, Methot)
- Recruitment lognormal bias adjustment protocol to obtain consistent results in Max Likelihood estimation (Methot and Taylor)
- Effect of spatial structure on performance of assessment models (various)
- Reports from the UW team to be presented today

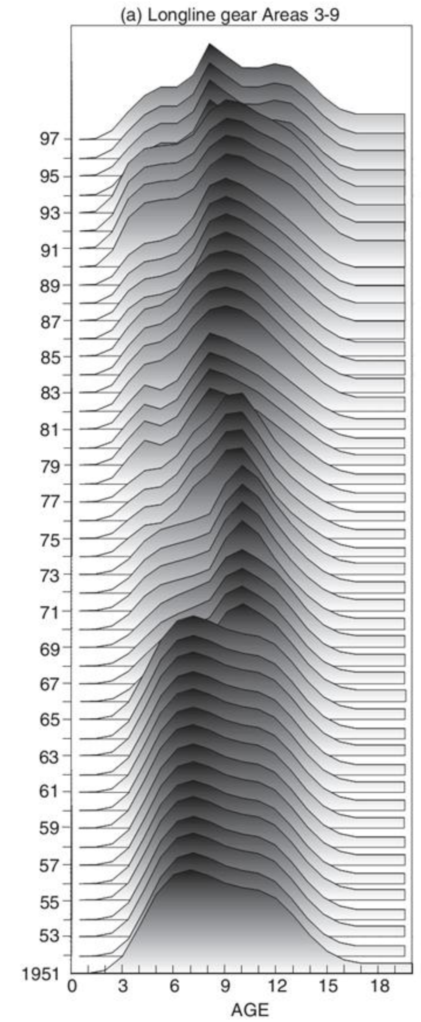
Parameter Priors and Linked Assessments

- Meta-analysis: Two recent papers by Thorson, Taylor, Stewart and Punt develop a mixed effects model to integrate results across SS applications for several west coast species
  - Estimate life history ratio: M/K
  - Estimate coherence in recruitment deviations
- Survey Q, F, survey process errors, and other factors are amenable to derivation of informative priors by linking assessments of multiple, co-occurring species

Are We Estimating the Right Factors?

Some Common Practices

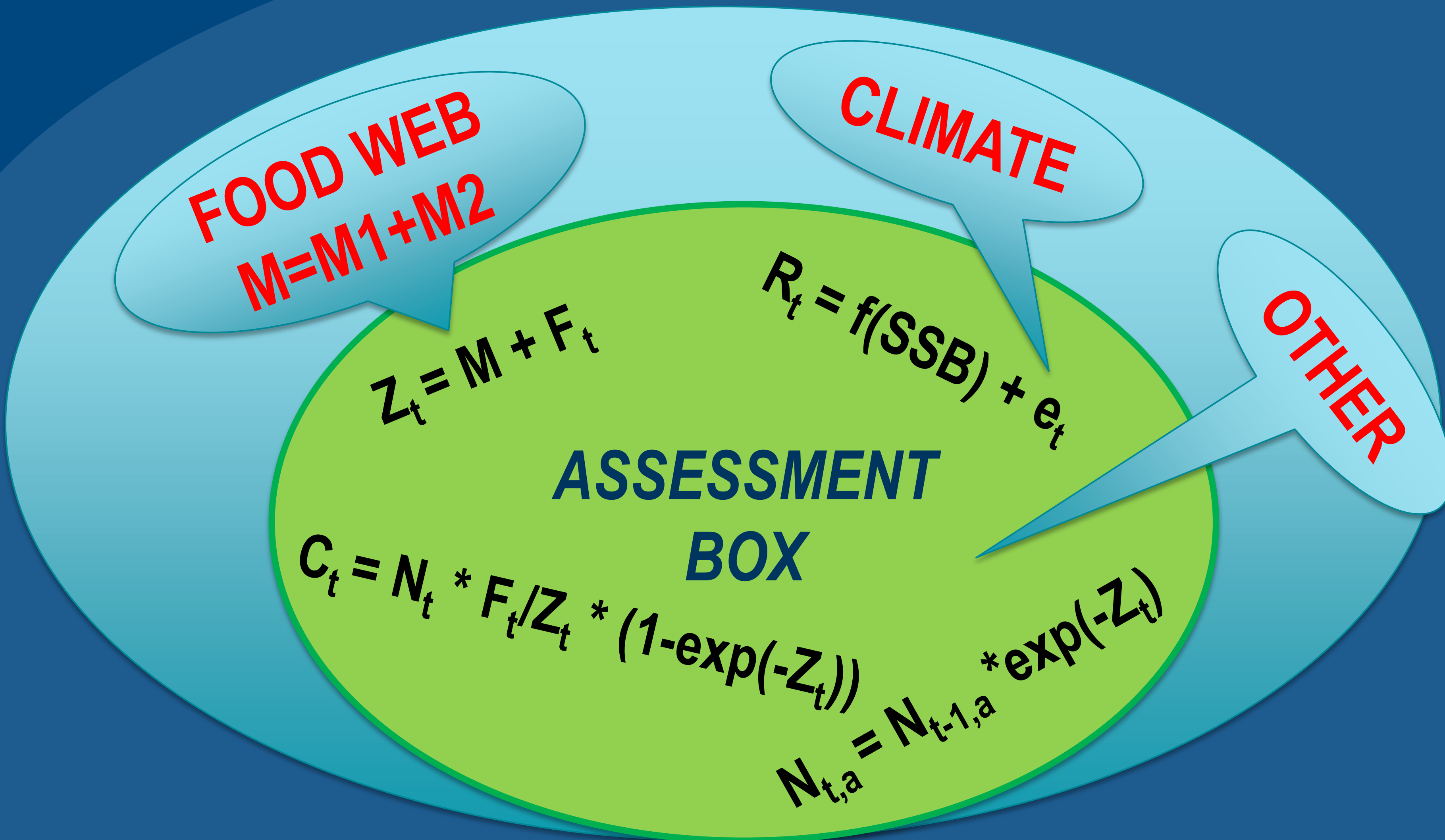
- Hold M constant, but contemporary M is among the least known factors!
- Put parametric, or complex non-parametric (right), statistical constraints on selectivity of fisheries
- Use age-specific surveys, so each has fully independent Q
- Treat survey Q's as having only uninformative priors
- Estimate population conditioned on above, but many degrees of freedom in the age composition data go into the selectivity estimation



What Could We Do Differently?

- Gear experiments, tagging studies and spatial distribution studies to make direct measurement of selectivity, or linkage of Q between ages in survey; include goodness of fit to selectivity data in models
- Gear experiments and and spatial distribution studies to put priors on overall survey Q
- With information on Q and selectivity; M estimation becomes more feasible

Ecosystem and Assessment Models



Three Approaches

1. Deterministic: Expand system so that **$M_t = f(E_t)$** is now inside the system
   - Multi-species models take this approach (Curti et al)
   - Also recruitment driven by environmental time series
2. Random Effects: Treat **M_t** as a random process and integrate over the range of possible values to obtain an estimate of the average performance of the system, and its variance. The posterior distribution of **M** is determined by the prior on **M** and the information in the conventional "inside the system" data. **E** remains outside the model system.
3. **E** as **DATA**, like a survey of the state variable **M**.

External Factors as Data Regarding Deviations

- Expected value of factor $\mathbf{E_t}$ is a function of state variable $\mathbf{M_t}$. Same logic as expected value of a survey is a function of the state variable $\mathbf{Biomass_t}$.
- Model includes the logL from deviations $\mathbf{(E_t - e(E_t))}$ in the objective function
- Example:
  - Recruitment as a random process with annual values $\mathbf{R_t}$
  - A survey, $\mathbf{O_t}$, of young fish is considered a measure, with sampling error, of $\mathbf{R_t}$, so $\mathbf{e(O_t)=f(R_t)}$
  - This survey could have been an annual measure of some environmental factor. From the assessment model's perspective it is just a datum that is informative about $\mathbf{R_t}$
  - The estimates of the $\mathbf{R_t}$ will depend upon the conventional data, e.g. age compositions and young fish surveys, ***and*** the new ecosystem/ environmental data
- Stock Synthesis provides this approach for the recruitment process, and soon other random processes

SUMMARY

- Generalized assessment models can provide consistent results from a diversity of data types
  - Need best practices guide and good A.I. in model interface
- Simulation studies are key to understanding model performance in face of diverse data and structural situations
  - Must build process error generation into these studies

LOOKING FORWARD

- Meta-analysis across species will improve informative priors
- Environmental data and ecosystem model outputs will routinely be used as “data” about time-varying model processes
- Direct studies on selectivity and catchability will provide better estimation of M and the population
- A protocol for consistent derivation of reference points and harvest policies when vital rates are time-varying or ecosystem linked, including detection of regime shifts, will be developed
- Models that include spatial sub-structure will be applied in relevant situations
- Perceived boundary between single species and multi-species models will disappear; just more code and more to review
- Assessment results are imprecise and will feed into MSE evaluated management procedure, not simple control rule: C=F*B .

NOAA FISHERIES

Grand Questions Confronting Assessment Scientists

Richard D. Methot Jr.

Science Advisor for Stock Assessments

Workshop on Stock Assessment Methods

Boston, MA

July 16, 2013

Grand Questions – V1

- Tier I
  - Utility of ageing data
  - Dome Selectivity
  - Change in selectivity over time
  - When is VPA approach better than SCAA?
- Tier II
  - Trends in M with age and time
  - Retrospective patterns
  - Role of contrast in time series
  - Performance of management procedures with biased indexes
- Tier III
  - Internal or external estimation of Spawn-Recruit
  - Age vs length vs stage composition
  - Catch error and discard estimation
  - Age reading errors

Grand Questions – Clustered

- Age
  - Utility of ageing data
  - Age vs length vs stage composition
  - Age reading errors
  - When is VPA approach better than SCAA?
- Contrast
  - Role of contrast in time series
  - Dome Selectivity (affects estimation of Z from composition data)
  - Change in selectivity over time (affects contrast in the age data)
  - Trends in M with age and time (confounds with selectivity)
  - Internal or external estimation of Spawn-Recruit
  - Catch error and discard estimation (affects stability provided by known catch)
- Performance Metric
  - Performance of management procedures with biased indexes
  - Retrospective patterns

Stock Assessment Goals

- What harvest policy is sustainable and provides balance between preventing overfishing and attaining maximum fishing opportunities?
- Does current level of fishing (F) exceed that policy?
- Has abundance (B) been so reduced by past fishing as to put the stock and ecosystem at risk?
- What future catch would implement the policy?

Needed

- Good practices guide on how to deal with assessment issues, for example:
  - When to not invoke domed or time-varying selectivity;
  - How to deal with factors (e.g. $M_{t,a}$) that are poorly informed by available data;
  - How to do model averaging across a broad range of model structural uncertainties?
- Simulation studies
  - Observation error only
  - Process error also

Strategic Initiative on Stock Assessment Methods (SISAM)

BOSTON

WCSAM

2013

WORLD CONFERENCE ON STOCK ASSESSMENT METHODS

WCSAM Workshop 15-16 July 2013, Boston USA

Topics for Discussion

- Guidelines for Stock Assessment
- Applying Multiple Models to Real Data
- Guidance on Future Simulation Testing
- Simulation ‘Self-Testing’
- ‘Cross-Model’ Simulation Testing
- Next Steps for the Strategic Initiative on Stock Assessment Methods

Guidelines for Stock Assessment

- Previous guidelines have been from leaders in the field (Beverton & Holt, Ricker, Hilborn & Walters, Quinn & Deriso) or chapter-authored books (e.g., Gulland), but those are now out-dated.
- Guidelines should identify validated approaches but allow for innovation.
- Consider guidance from traditional statistics (inspection of residuals) and recent advances in model building as a starting point for guidance on stock assessment modeling.
- Guidance extends to data (CPUE, surveys, …) and process (documentation, peer review, ..)
- Guidance needed on determining the ’best model.’

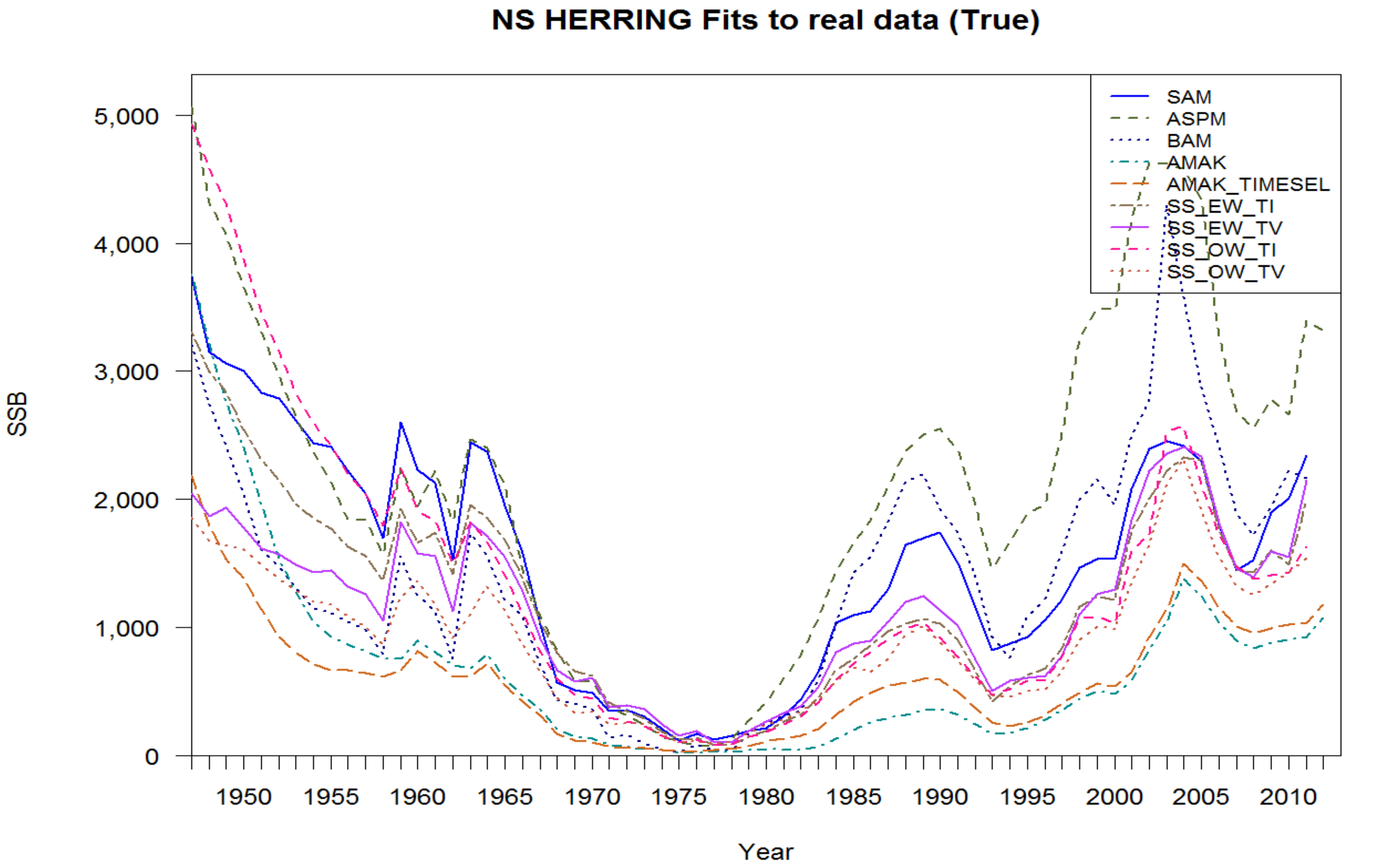
Applying Multiple Models

- Terminology should be defined (‘model’, ‘configuration’, ‘software’, ‘framework’, ‘mechanisms’, ‘process error’) if not standardized.
- The most appropriate model category depends on the type of data available.
- The most appropriate model configuration within a model category depends on the information content of data.
- Applying multiple models or multiple model configurations helps to evaluate model uncertainty but can present challenges for management advice.
- Applying multiple models to real data has limited information on model performance (we need to know the "truth" to evaluate model performance)

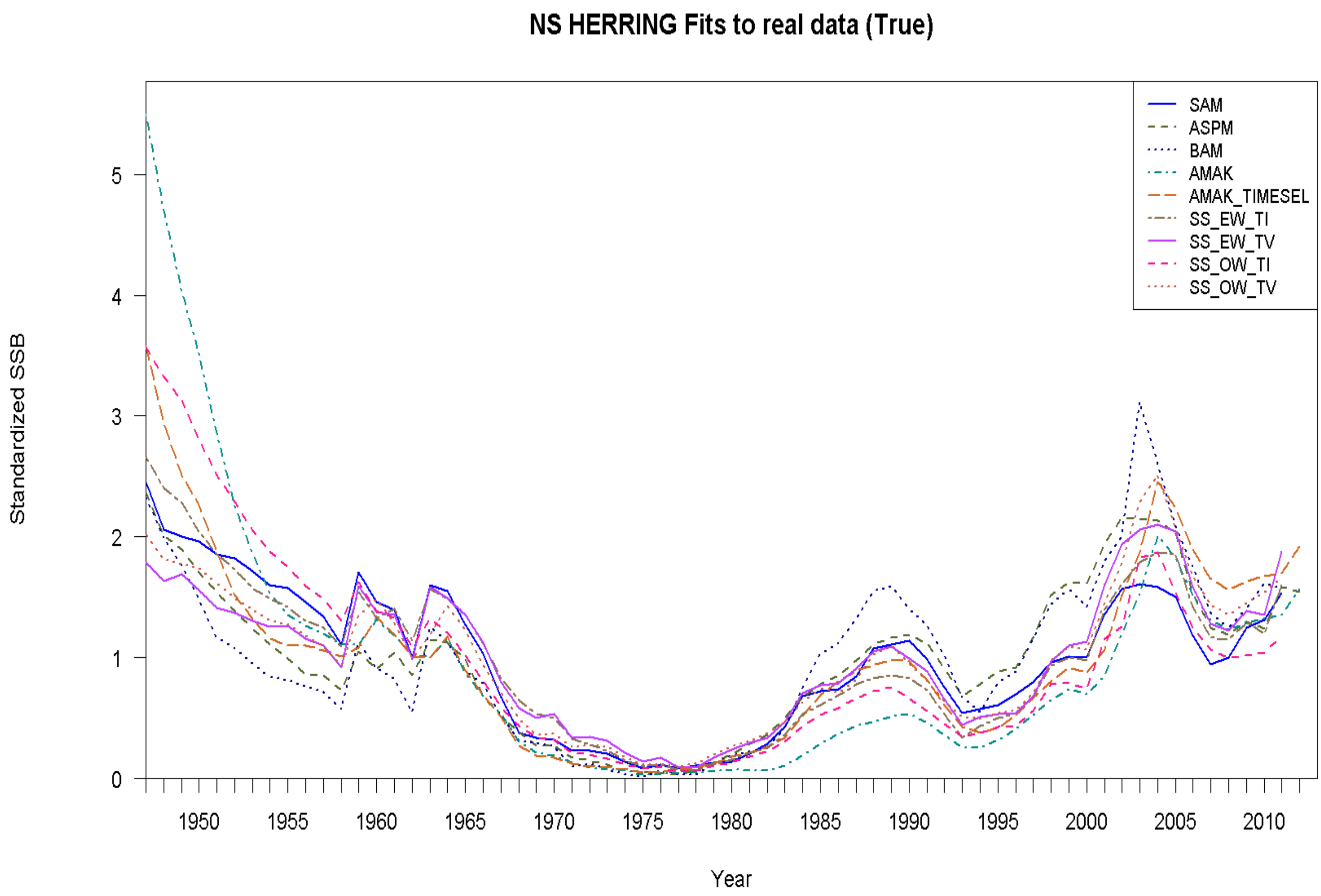
Guidance on Future Simulation Testing

- Simulations should be designed to meet defined objectives:
  - Generic guidance, specific validation vs. specific problem solving
  - Degree of operating model complexity (single-species, sexual dimorphism, ecosystem, spatial patterns, alternative population regulation, fleet behavior, …)
  - Data availability (and time series length)
    - Conventional (fishery catch, CPUE, surveys, length/age compositions)
    - Non-conventional (tagging, consumption, surveillance, …)
  - Performance criteria; e.g.,:
    - Recent stock size and fishing mortality (absolute, relative, retrospective adjustment?)
    - Stock status relative to reference points (absolute or relative?)
    - Short-term catch forecasts,
    - Uncertainty in short-term catch forecast
    - Medium-term projections (e.g., rebuilding plans)
    - … (see Marianne Robert et al., 2010)
  - Integration of uncertainty? Bayesian?

Absolute SSB



Relative SSB



Simulation ‘Self-Testing’

- Self-testing (same operating model used to simulate data as the estimation model applied to simulated data) is essentially a consistency check.
- When ‘significant’ differences are found (e.g., ‘truth’ is a low probability among estimated realizations), the inconsistency should be investigated:
  - Audit code to confirm that the operating model and estimation model are identical in all settings.
  - Apply estimation model to ‘perfect data’ (e.g., CV~0, N=high) to identify source of inconsistency (observation error or structural inconsistency)

‘Cross-Model’ Simulation Testing

- A matrix of ‘Cross-Model’ comparisons (different model used as the operating model to generate data than the estimation model applied to simulated data) is informative, even to communicate which combinations were tested.
- Exploring scenarios (model configurations) in a common software is more informative than comparing different software packages that have the same configuration.
- In the real world, attempts would be made to investigate and reconcile divergent results.
- When different models produce divergent results when applied to simulated data, the source of divergence should be investigated:
  - Adding penalties to the estimation model for deviating from the ‘truth’ generated from another operating model and inspecting likelihood components may identify the source of the divergence.

Next Steps for SISAM?

- Near-Term (Thursday afternoon session of WCSAM)
  - Present simulation approach (Doug Butterworth)
  - Present conclusions from simulations (Jon Deroba)
  - Present discussion on 'grand questions' (Rick Methot)
- Longer-Term Action Plan (Steering Group – Friday lunch)
  - Results from simulation exercises suggest that significant investment is needed to address model uncertainties.
  - Refined simulation exercises
    - Make sure simulations produce all of the data needed for each model category.
    - Promote ease of participation (common data formats, etc.)
  - Development of 'good practices' guidelines
  - Topical Workshops (e.g., CAPAM selectivity workshop)