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

- **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

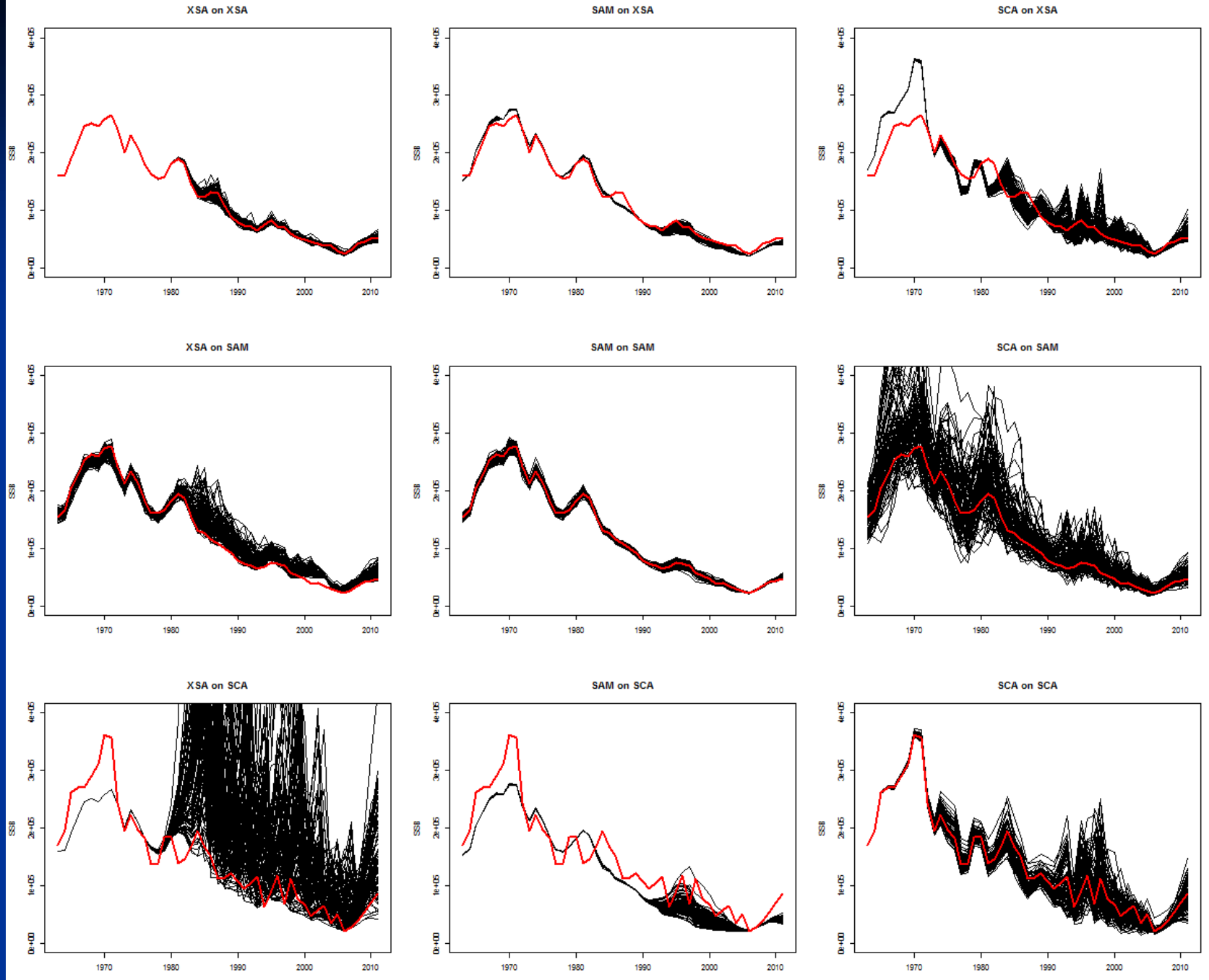
PERFORMANCE COMPARISON PLOT

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

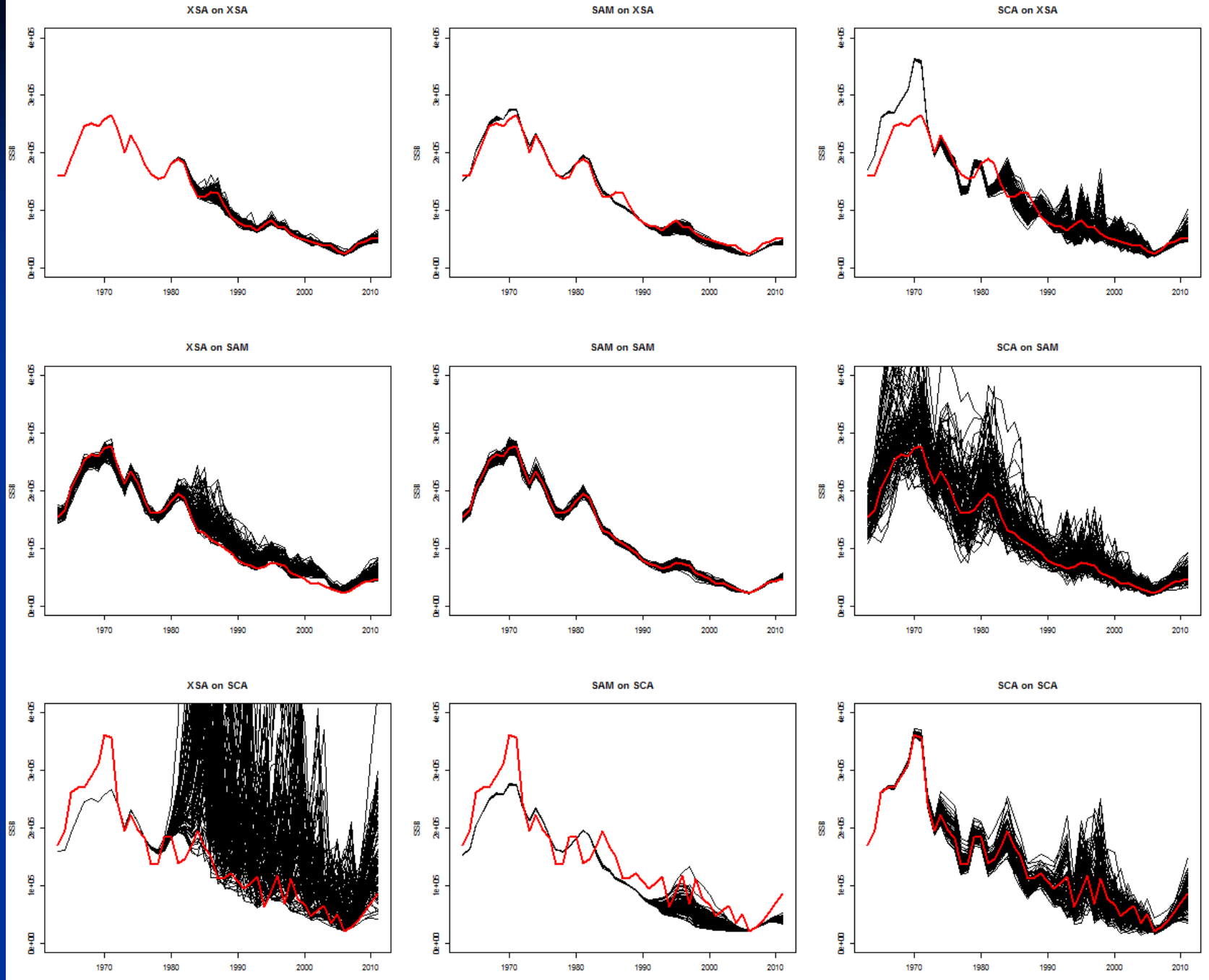
Cell contents: Performance statistic, here SSB

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
0e+00
1e+05
2e+05
3e+05
4e+05
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