

# Performance Testing of Indicators: From telling stories to informing decisions

Jake Rice  
DFO-Canada

# A Step beyond Describing Pattern

- This morning we heard state of the art on using ecosystem indicators to describe ecosystem status
  - INDICATORS ARE USEFUL (with qualifications)
    - NOT a substitute for expert knowledge
    - Challenge to move to real-time reporting
- Now what do we expect decision-makers to DO with this information if we do provide it?
  - Can't get away with the case-by-case expert "excuses"
  - Can't get away with asymmetries in interpretation,
  - Must live with "decide now – correct later" reality

# What we must believe about Indicator Performance if we are promoting them

That Ecosystem Indicators can be used in existing (or new) management frameworks to help:

- improve sustainability of human activities in the sea;
- reduce ecosystem effects of the human activities;
- restore the 'health' of ecosystems and populations;

The ways that they can contribute include by helping to:

- Set BETTER targets and limits for traditional indicators
- Allow evaluating status of DIFFERENT ecosystem properties relative to their targets and limits
- Make better CONTROL RULES for managing towards objectives

# What is meant by **Performance** of Indicators - and who is interested?

How well do indicators...

1. represent the true properties of the ecosystems they are meant to measure? – (largely SCIENCE concern)
2. track progress towards meeting management objectives: ecological, social, economic?- (Science and policy concern)
3. respond to changes in management measures ? (Policy and Science concern)
4. **inform the decision making process** in management? (Largely POLICY concern)

# What APPROACHES can be used for performance testing

Analytical approaches for the OUTCOMES of:

- 1 - capturing true dynamics: retrospective modeling and analysis of historical data.
- 2 - achieving management objectives: 1) + scenario modeling with management interventions) –
- 3 - evaluating effectiveness of management tools:  
1) + 2) + Full Management Strategy evaluations.

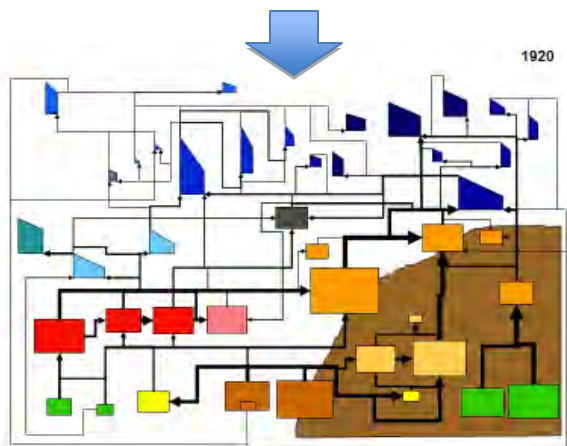
However must add **formal decision analysis** for

4) Testing and refining control / decision rules

BRIEF look at Simulations – familiar and other talks.

Focus will be on DECISION ANALYSIS tools

# SIMULATION of performance relative to 'true' attributes



Observation process & error

indicators

Some of indicators that performed well relative to true system attributes:

Biomass of groups

Pelagic: Demersal

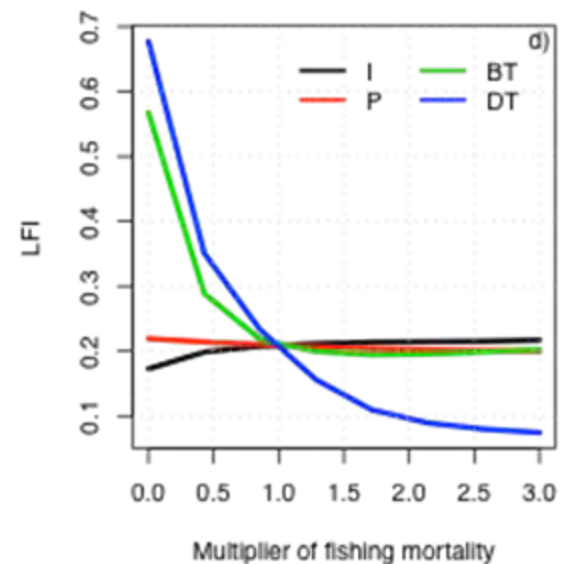
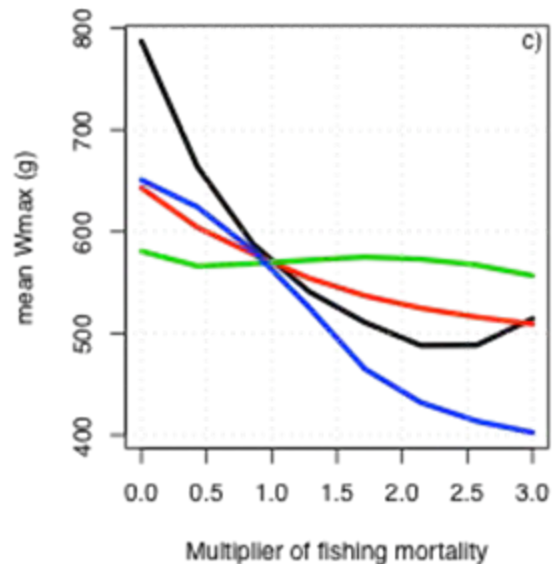
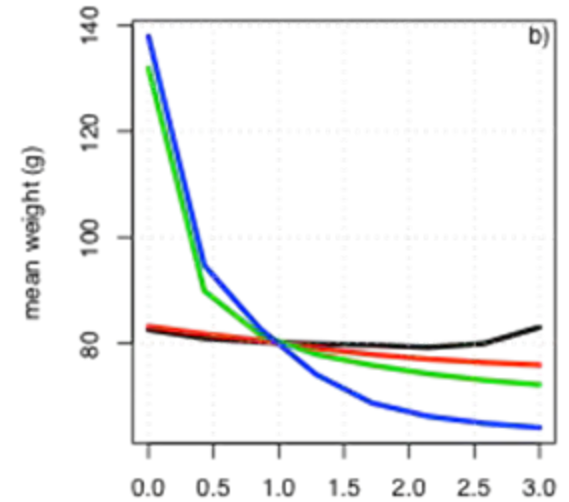
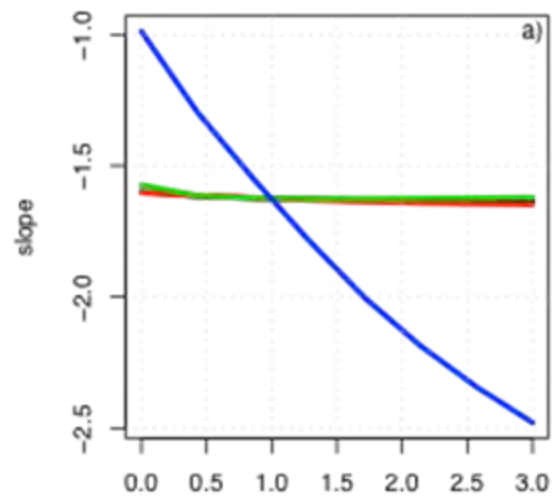
Piscivore: Planktivore

Biomass size spectrum slope

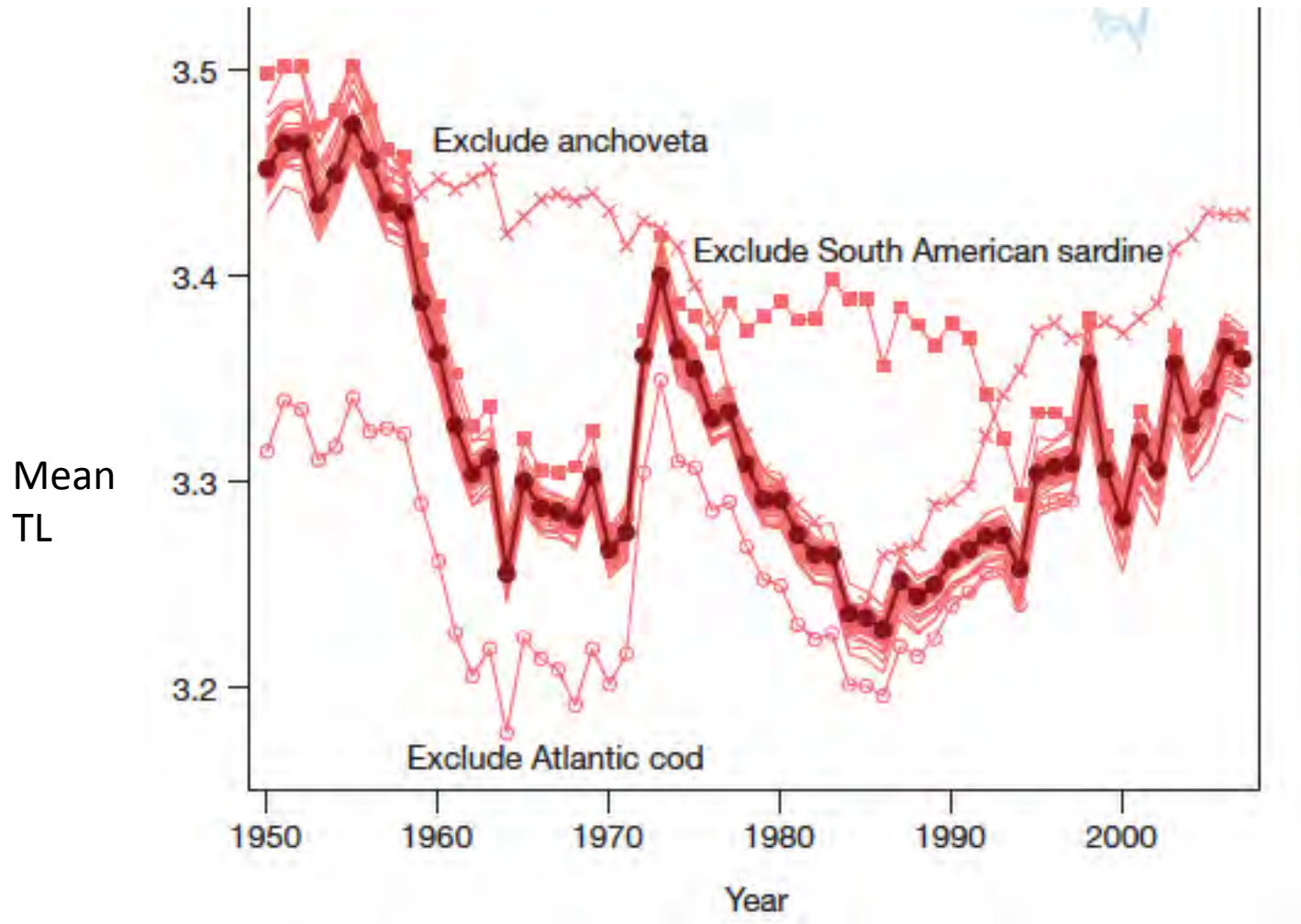
# OTHER USES start to Deal with: Sensitivity to functional form of pressure – state relationship

Indicators respond differently depending on exploitation pattern of various fleets [Industrial pelagic trawl beam trawl demersal trawl]

Blanchard unpb.



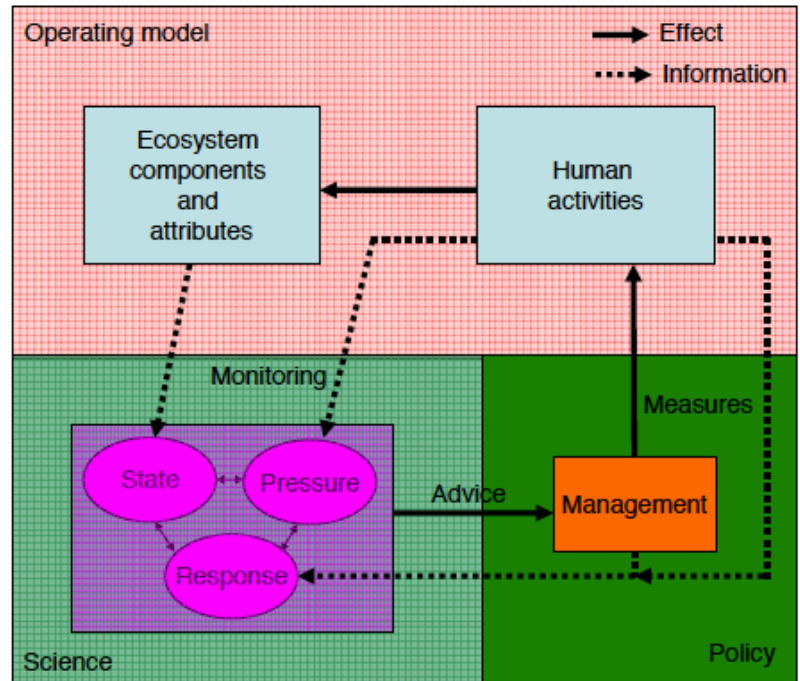
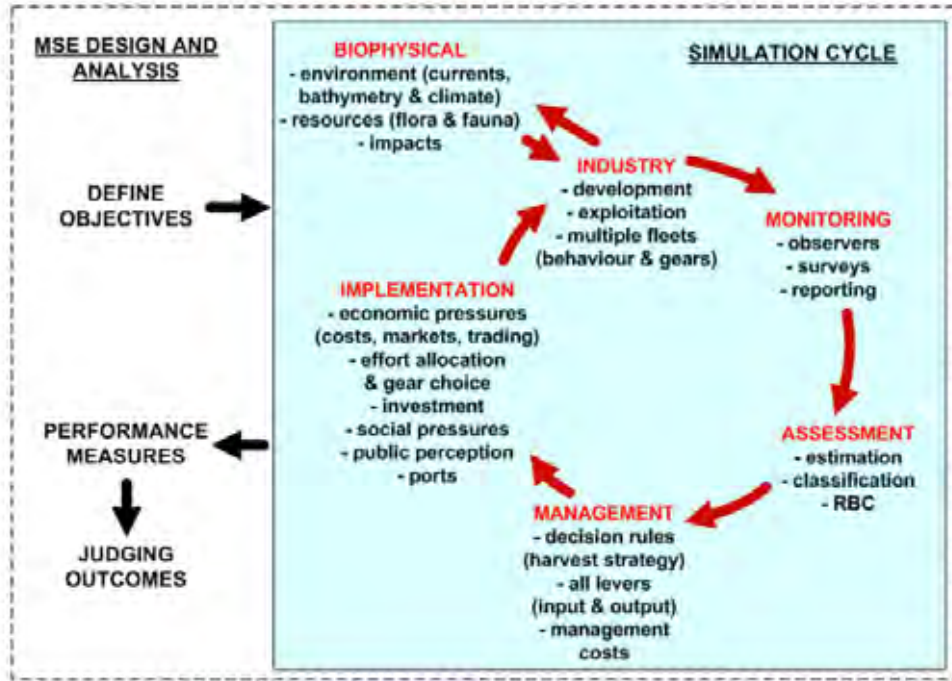
# and Sensitivity to taxonomic bias (etc)



Can be explored with data and models



# Management Strategy Evaluation

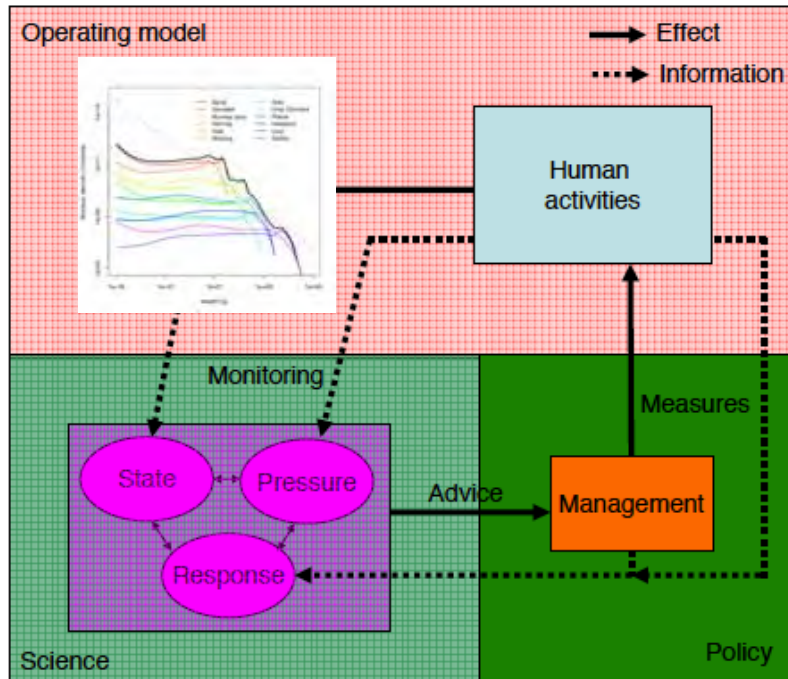


Fulton et al 2007

Piet et al. EU F6 project "IMAGE"

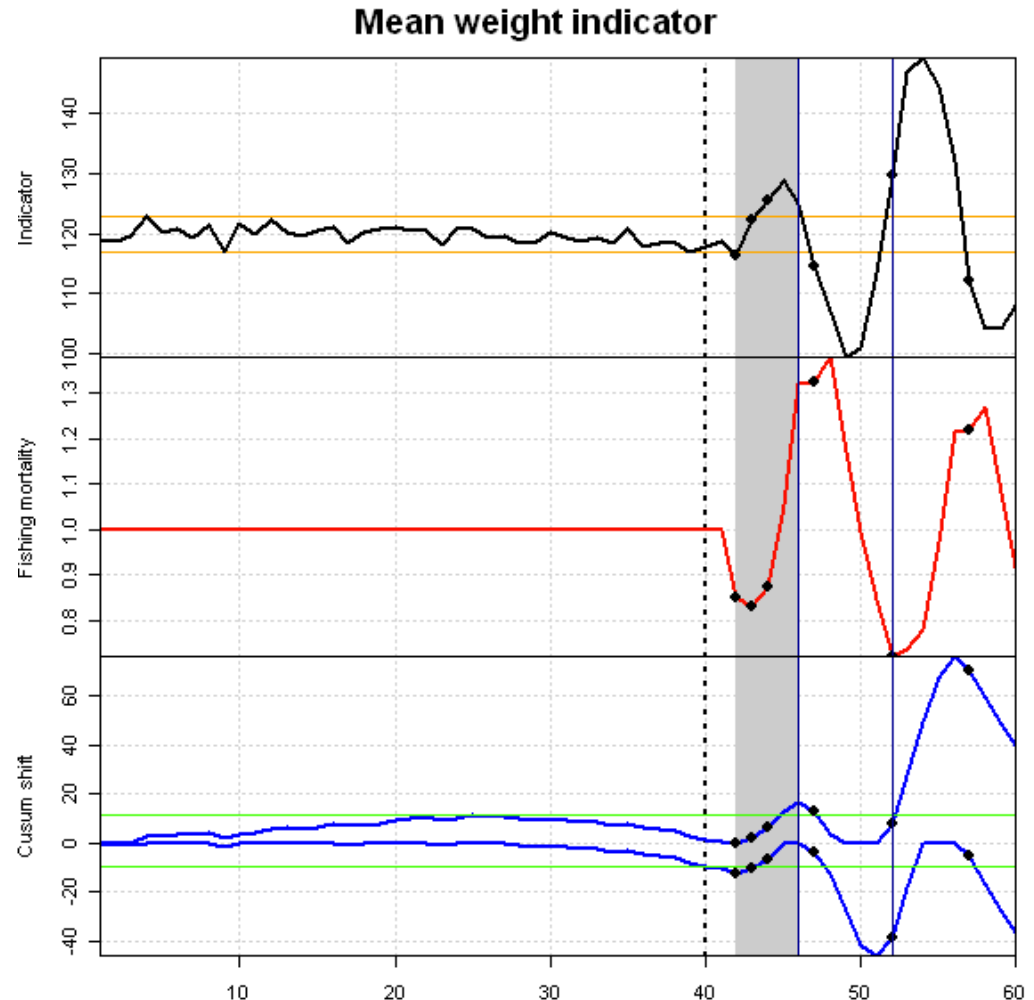
Which indicators to use in: monitoring, assessment, performance measures (combined with targets) for meeting objectives?

# Example of responsiveness to management



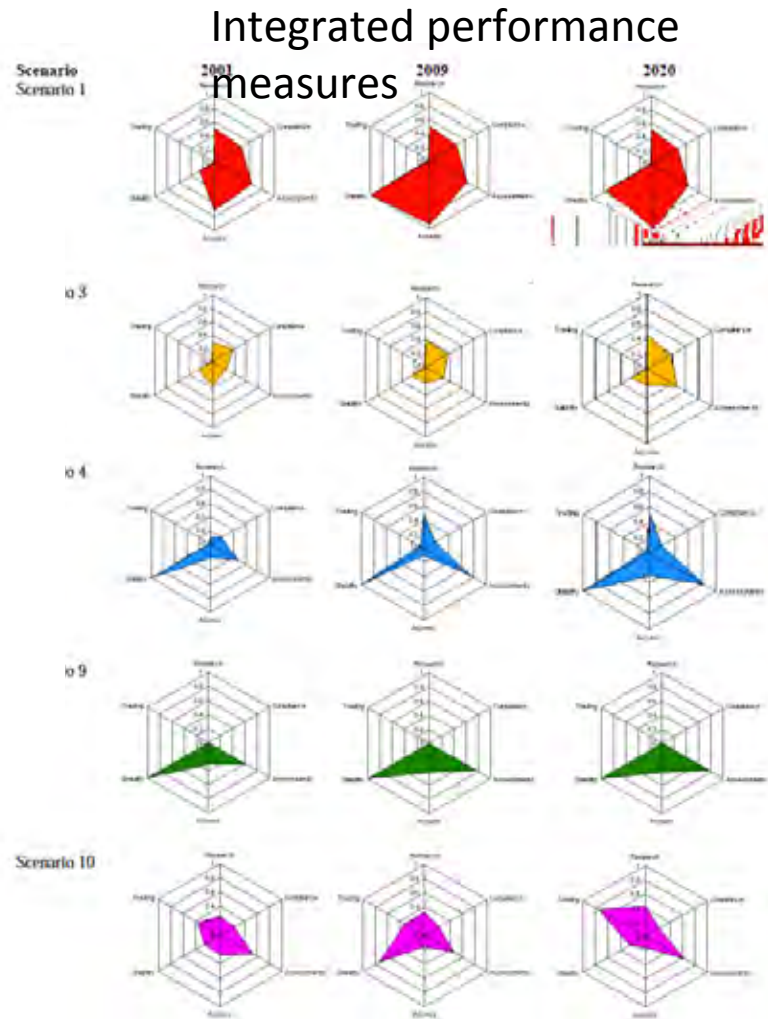
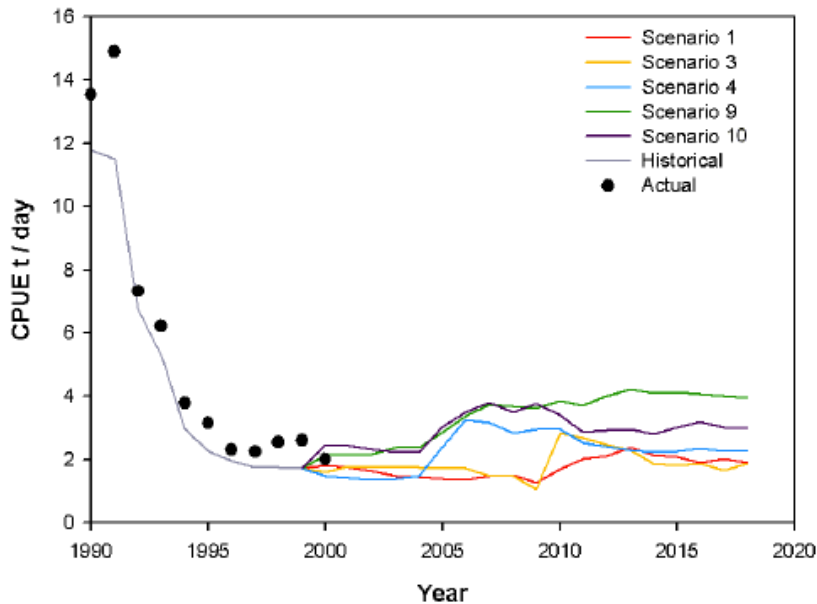
Operating model: multispecies dynamic size-structured model of North Sea + observation process and error

Test of simple harvest control rule (CUSUM, Mesnil et al 2009) based on indicator



- Indicator
- Range indicator
- Indicator decision points
- Fishing mortality
- Cusum indicator
- Range Cusum
- Cusum decision points
- ... End of reference period
- █ Indicator response period

# Example of retrospective modeling & management strategy evaluation



# Performance testing of Ecosystem Indicators

## DECISION RULES

- Focus will be on Retrospective Decision Analysis to improve decision rules used in Policy and Management
- **Value in supporting decision-making**
  - When the work really starts to be taken seriously because it is USED in policy and management
  - Simulations have become very complex and messages hard to extract and communicate
  - **Requires one to think about some different things, and some familiar things differently**

# Think about WHAT?

Think about some different things:

Can a policy / management decision use this indicator directly (in a decision rule), indirectly (determine which rule to use), or not at all? (Jason's book}

Had the rule been used in "last year's" decision, what would the decision have been? Would it have been "better" than the one that was actually made?

Think about some things differently:

"Processes" are background not focus of the analysis

Will the indicator be available in time for action to be taken

Do the reference points lead to optimal decisions? (What IS an *optimal* decision?)

# Hinges on knowing the policy / management question (s)

- “The question” is not about the ecological things the scientists find most interesting
- “The question” is about the decision to be made.
  - STRATEGIC: What target harvest rate to apply
  - TACTICAL: What size harvest to allow this year
- The role of the indicator in the decision is phrased like: If the indicator has value ‘X’ then take action ‘Y’
  - Action can be binary – Intervene or not
  - or continuous – adjust a control (e.g quota) up or down

# Simulation testing still in the picture

- Need to project something forward in time from specified state, under specified conditions.
  - “Something” depends on what the mgmt decision addresses.
    - Often just a single population or small number of interacting populations and properties
  - “Specified state” is conditions at the time decision was made
    - Usually a populations size and indicator value
  - The “specified conditions” are whatever the decision said should be done (Setting quote in example)

# The simulation task is often (not always) simpler than full MSE

**Rarely need to model whole ecosystem. Focus on-**

- first-order properties affected by decision,
- the properties constituting the management decision rule
- may be use-dependent or use-independent properties, depending on the decision rule

**Projections do not need to be as far into the future.**

- Often focus is historical period of past decisions
- Into future the decision rule feedbacks often dominate over system interactions
- Can add regime-type interventions arbitrarily as test



# Approach by Retrospective Testing of Historical Decisions

- What decision *would have been made* had the indicator been used in decision-making in the past. (Often data-based not model-based)
- Contrast with decisions actually made to see incremental benefit of the indicator
- Benchmark for comparison depends on
  - What the decision is (example – fisheries quota )
  - Whether historical decisions followed advice given at the time.

# Necessary Pieces for the Evaluation

1. Time series of the ecosystem indicators of concern, for selected historical period
2. Time series of the actual status of the ecosystem property represented by the indicator over the time period in 1.
3. A candidate rule (or rule-complex) for using the candidate indicators in 1 in decision-making.
4. A standard for judging “performance” of the rules in (3) using the indicators (1)

# 1. Time series of values of the ecosystem indicators

- Choice of indicator (s) is up to the researcher
  - Need plausible reason to think the indicator reflects a relevant ecosystem property:
    - Impacts the dynamics of the harvested stocks
    - Will be impacted by changes in harvested stocks
  - Needs time series of ecosystem indicator to overlap with time series of status indicators of ecological property of interest (the “stock”)
  - Source can be data-based or model-based

## 2. Time series of the actual status of the ecosystem properties of interest

Fisheries - Use the standard indicators of status of the stock and fishery, typically SSB and F

- The annual estimates of SSB and F from “best” assessment are typical basis for management decisions
- Converged SPA properties mean one can know the “right” answer (at least more accurate answer) for the past
- Can be augmented with other stock characteristics (age, size composition, stock structure)
- Can be augmented with more inclusive ecosystem features (these usually *incremental* not replacement)
- If the “real” property is abstract or synthetic, need data or model to reconstruct past “true” status.
- May only be able to smooth a noisy indicator



# Time series of the stocks and fishery - Complications

- When basis for management has changed over time, the “right” benchmark for comparison will have changed
  - Especially true with implementation of new management tools like spatial controls
- Role for simulations in what “what would have been stock history had regime X been in place for whole time”
  - COMPARATIVE evaluations often need only 1<sup>st</sup> order effects
- Reconstruction of historical record of what stock trajectories and management decisions would have in the past much simpler than projection of stock trajectories and linked decisions in the future.

**We will always have more information about the past than future, whether fisheries or any other decision**

# 3. Candidate Decision Rule

- What the decision-maker should do, given the value of the indicator
- Key roles for Limit, Precautionary, and Target reference points (RPs)
  - Keep risk of falling outside the Limit RP very low
  - Use Precautionary RPs to take account of ALL sources of uncertainty and keep management decisions from being “all on / all off”
  - Targets guide choices IF within precautionary boundaries.

# Making the Decision rules

- Classic – Must specify Reference Points and rule:
  - Outcome of management is to avoid ( $B_{lim}$  and  $F_{lim}$ )
  - Set risk tolerance of approaching outcome ( $B_{pa}$  &  $F_{pa}$ )
  - Specify rule for how harvest varies with position relative to the reference points
- Can make rules for reference points relative to fishery decisions just as readily –
  - empirically - historical time series for GAM-like work
  - Theoretically from modelling
  - Can be bi-variate continuous, conditional (if-then), Boolean (multiple T/F) or other
- WITHOUT A RULE CAN'T USE THE INDICATOR



# 4. Evaluation Framework allowing tests of three Levels of “RELIABILITY”

1. Is the ecosystem indicator a reliable signal of the ecosystem property of the indicator itself? (Usual standard)
  - Least useful to management,
  - Circular unless there is a TRUE time series for the ecosystem property (like “converged SPA”)
2. Is the ecosystem indicator by itself a reliable signal of a more important ecosystem property (the “stock”)
  - Helpful in data poor or high cost contexts
3. Does the ecosystem indicator, alone or combined with SSB (or a relative) produce a **reliable way to make decisions** for the stock (s) being harvested (ecosystem being used)?

# Signal detection theory – How well does STOCK indicator reflect STOCK dynamics

Actual Trajectory of ecosystem property (Stock )	Conclusion based on Indicator (annual SSB estimate)	
	Property (Stock) is getting better	Property (Stock) is getting worse
Getting Better	Hit	False Alarm
Getting worse	Miss	True Negative

# Standards for Evaluating Performance

- Perfect indicator causes no misses and false alarms
- Systems with noisy indicators, stochasticity, or imperfect link of indicator to ecosystem property will produce the two types of errors
- Series of similar errors inform us that there is a problem with the signal and/or decision-rule for using it
- Binary choice can be relative to benchmarks other than “no change” – whatever causes management to act or not act
- Framework can be made probabilistic, at modest additional computational cost and often not much gain in insight
  - Primary Goal usually to get rid of errors,
  - Matching scale of hits and misses secondary concern once a really reliable signal is identified

## C) Signal detection theory – HOW GOOD IS THE MANAGEMENT DECISION

The right decision for management of the stock	What the indicator- <b>rule</b> says management should do	
	Increase harvest	Decrease harvest
Increase harvest	Correct action	False Alarm
Decrease harvest	Miss	Correct Action

# What Signal Detection Theory Brings to Comparative Testing

Focuses on the TYPES of errors as well as RATES

- The **costs** of the two errors may not be the same
- **Risk tolerances** of the costs differ among participants in the decision-making processes
- Decision rules can be tuned to produce balance of errors that meets management objectives / risk tolerances
  - Few statistical optimization algorithms differentiate the types of errors, but management does
- Framework places dialogue into P (miss) and P (False Alarm) and can be used in all sorts of probabilistic modelling

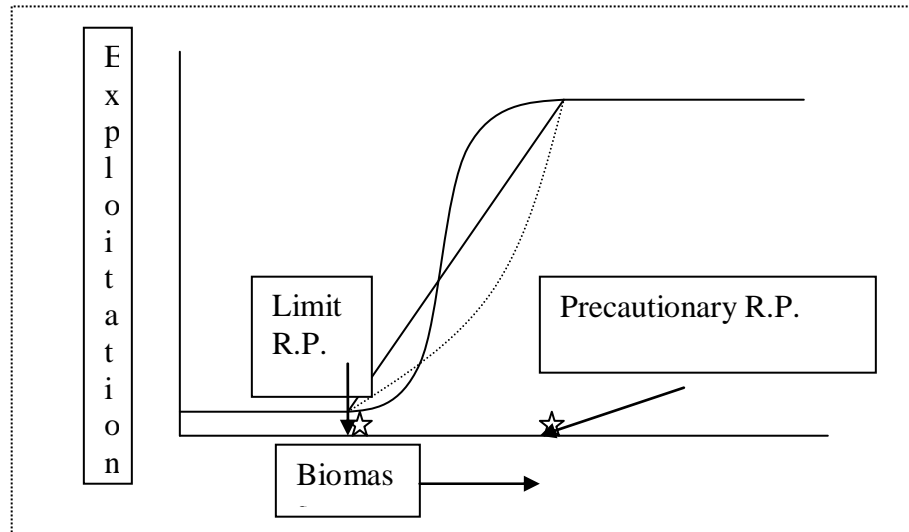
# Step-wise preparatory work for the SDT Evaluation Process

1. Look up the “**Correct**” decision vector  
from “converged” SPA or best use of hindsight (model or expert)
2. Look up the **Actual** decision vector for the current rule  
Simply the record of whatever was actually done
3. Evaluate the outcome matrix from SDT and establish the  
error profile for the base decision rule
4. If management has changed: “calibrate” the actual decision  
vector and the “correct” decision vector for consistency
  - What would the decisions have been if one (current) management  
approach had been in place over full time series? SIMULATION
5. If Management has changed evaluate the outcome matrix  
of the calibrated decision vector (4)

# How to Compare New Rule with Ecosystem indicator - Stepwise

- A. Specify the rule to use the ecosystem indicator (EI)
- Quota =  $f [(EI_{\text{current}} - EI_{\text{ref point}}); \text{risk tolerance}]$  OR direct use of Ecosystem Indicator in management
  - Quota =  $f [(B_{\text{cur}} - B_{\text{lim}}; \text{risk tol.}, f [(EI_{\text{cur}} - EI_{\text{ref point}})];$   
adjust standard management rule by ecosystem status ,
  - the  $EI_{\text{ref point}}$  can be a “limit”, the mean value of the EI, or any ecologically justified trigger
  - the function can be continuous, conditional, Boolean threshold, or anything else.
  - The EI rule can be to change the standard ref points  
 $[B_{\text{ref points}} = f(EI_{\text{current}} ; EI_{\text{ref point}})]$ 
    - Function can be continuous, conditional , etc

# Types of control rules





# Using the rule to produce the ecosystem decision vector

- B. Simply apply the rule in A to the annual estimates of the EI (from the historical time series) and if needed the annual estimates of biomass (or F) from the stock.
- C. Calculate the SDT outcome matrix comparing the “Ecosystem Decision Vector” to the “Correct” outcome vector.
- D. Compare the outcome of C to the outcome of step 3 of the “non-ecosystem” decision rule

Keep the comparisons “fair”: The only difference between the “ecosystem decision vector” and the vector to which it is compared should be the contribution of the ecosystem indicator(s) to the values in the decision vector

# Tuning the Decision Rule

- D. Compare matrix from C to matrix from 3 (OR 5)
- Use SDT outcomes as basis for improving rule parameters and performance
  - Can move  $EI_{\text{ref point}}$  or try a different function of proportionality for EI relative to “SSB” to get the distribution of Misses and False Alarms that reflects management/policy values
  - Algorithms exist but trial and error fast and generates important insights

**DOES INCLUDING THE ECOSYSTEM INDICATOR PRODUCE BETTER OUTCOMES?**

# Necessary steps for testing indicators for management

- 1) Identify objectives and RPs to be used to meet them
  - Use policies, reference levels identified appropriately, or trend-based approaches to establish targets
- 2) Design decision rules that are linked to management control measures (quotas, effort, area based management)
  - Need evidence that a change in control would cause a change in the indicator
  - Identify acceptable level of risk,
- 3) Select method for assessing performance (simulation, signal detection theory)

# Which indicators should be used in management?

Screening candidate indicators:

- Are they CREDIBLE (data and simulations)
- Are they LEGITIMATE (social sciences)
- Are they RELEVANT
  - Is there credible link to the objective(s)?
  - Does a change in the control measure cause a change in the indicator as expected?
- Evaluate Performance
- Choose SUITE of known net performance

ADVICE is expert narrative backed up by indicators

- “Don’t shoot the pilot just because you have instruments in the cockpit!

WITH THE RIGHT HELP AND THE RIGHT  
TRAINING WE CAN DO ALL THIS

