Comments on the SAFE *Ecosystem Considerations* appendix and the PICES North Pacific Ecosystem Status report, and review of Day 1

Jake Rice

Fisheries and Oceans Canada, 200 Kent Street, Stn. 12S015, Ottawa, ON, K1A 0E6, Canada
E-mail: ricej@dfo-mpo.gc.ca

My comments arise from reading the SAFE *Ecosystem Considerations* appendix and the PICES North Pacific Ecosystem Status report, all the background papers, listening to the first day's presentations, and to the discussions that took place between the presentations. These are placed within the context of my experience in several other domestic (Canada) and international (ICES, EU, FAO, *etc.*) fora. My comments are structured as follows. First, I looked at the overall messages from the reports. Then I tried to assess what was missing, present but vague, or present but requiring greater discussion. Finally, I provide my own ideas of useful ways ahead.

My first observation is that you are in pretty good shape. A considerable amount of effort has been spent on objectives where there is an appreciation of the need for specificity, noting that objectives are converging from many sources. There is recognition of the need for socio-economic objectives and their differences from ecological objectives, on matching indicators to objectives, acknowledging two modes of use, and especially, there is no indication that the region is in desperation mode.

Both of the major ecosystem reports are very good, but I note that their different audiences are matched by their different content. Both reports have enough detail to allow users with preconceptions to guide the selection of the content. My suggestion for improvement is to avoid including details in the report that a reader/user of the report will not want. You should aim for a guidebook rather than an encyclopaedia and try to motivate and guide readers. Finally, make the big messages clearer.

Features that were either missing or under-represented in the papers and talks include the following:

- The DPSIR (Driver, Pressure, State, Impact, Response) structure has proven useful for organizing dialogue and in reducing numbers of indicators, and for matching indicators to their use in the overall processes.
- There was an overall absence of a risk management framework in the papers. The Fulton presentation has demonstrated one way for making progress on this topic. There is a need to focus more on displaying uncertainty. Of the suites of indicators, spatial content was missing everywhere, and I noted that size-based indicators are under-represented relative to their performance elsewhere (especially ICES).
- There does not appear to be a formal indicator selection process.

Several facets of the indicator issue were present but vague.

- There was no discussion of how to test the performance of indicators during the selection process.
  - NOT the same for indictors used in AUDIT function and indicators used in CONTROL function
  - AUDIT – Targets primary, limits secondary
  - CONTROL – Limits primary, Targets secondary
  - METHODS EXIST FOR TESTNG BOTH

- Where do we get the reference points?
  - Differentiate Indicator (say, SSB) from Reference Points (Bmsy, B35%, *etc.*)
  - Reversibility of impact? Responsiveness to management at all?

- The "classic" three-stage model (discussed below) should have ONE (NOT two) biological (or socio-economic) fixed points and the rest is making uncertainty explicit.

There are several areas that could use more critical thinking. These include when and how to use absolute-scale indicators *versus* relative-scale indicators. The experience with IUCN decline criterion for marine species is a case in point. Should there be different reference points for different regimes? Perhaps for the population size, NO, but for the uses of populations, YES, especially if likelihood of prompt detection of regime change is low. If the "traffic light" style of presentation is preferred, then the biological calibration of the cut-points is a crucial research topic, as are strategies for dealing with redundancies among indicators, and weighting of indicators when providing support for decisions. There is a need to understand what to do with tough decisions and multiple indicators that might reflect opposing trends. An example of this is the EU experience using just B and F (biomass and fishing mortality). I note that U.S. legislation on over-fishing and over-fished will not transfer readily to ecosystem metrics.

I was stimulated to ask what other field of science works with indicators in a similar context? My experience in psychometric research has some similarities. The fundamental underlying processes are critically important but they are NOT accessible to direct measurement. Therefore, indirect indices have proliferated and they are flexible, and are easily adopted. In psychometrics, "normal" is not a fixed point on ANY scale, but is a general "neighbourhood" in the centriod of the multi-dimensional space of the indicators. Usage focuses not on how close an individual subject is to exact centre of the neighbour, but rather on whether an individual subject is deviant in some particular direction and if so, by how much and what might be done about it. In psychometrics, a lot hinges on decisions based on the indicators and the ability to abuse and/or misinterpret indices is relatively easy. Hence the field has developed quite explicit and detailed guidelines for their use.

An important step is the selection process when indicators are being used in their Control function. A process which I find particularly attractive is derived from signal detection theory and its application to Human Factors Research. It has a 70-year history with its first flowering during WWII. It requires reconstructions of historical time series of indicator values. Once this has been developed, the next step is to reconstruct what a history of good decisions would have been. If it is not possible to do this retrospectively, how can we support any decision-making into the future, based on these indicators. Four outcomes are possible in evaluating an indicator:

1. HIT (something should have been done and the indicator said DO IT),
2. TRUE NEGATIVE (no management response was needed and the indicator said *status quo* OK),
3. MISS (something should have been done but the indicator did not say action was needed),
4. FALSE ALARM (nothing needed to be done but the indicator called for management intervention).

The results of this exercise can be represented in a 2×2 table. A perfect indicator has no Misses or False Alarms. The approach explicitly acknowledges that the costs associated with Misses and False Alarms are not the same. The approach allows users to choose a decision point on an indicator ("reference point") that minimizes the overall error rate or, controls the ratio of Misses and False Alarms that reflect their relative costs (*e.g.*, medical situations). In this way, it becomes easy to compare the performance of indicators.

Considering the Audit function in psychometrics, the diagnostics have a history of over a century of use. Many mistakes (and advocacy abuses) have occurred but many lessons were learned. The uses are numerous, from career aptitude testing, to assessing legal competency for actions, and diagnosing personality disorders. Their application has involved extensive validation testing and codification of professional standards.

The general approach is to have a large battery of "questions" (= "suites of indicators") – Binet, MMPI, Rorschach, *etc*., then test a large populace with the battery of questions. In addition to a

large number of subjects chosen at random from the general population, there is a special role for test sets, which involve individuals that are known with confidence to have specific disorders. The diagnostic tools are developed by determining a combination of weighted questions that group subjects known to share a specific pathology as distinct, while leaving most of the populace in a central cloud. To my knowledge, this approach has not been tried in ecology.

In the classic 3-stage mode (Fig. 18) for using indicators and reference points to guide decision-making, there is one fixed reference point. This is determined by some government responsibility – law or policy. The objective is typically to prevent any "serious or irreversible harm" (language of the Precautionary Approach from Agenda 21 of Rio). The best biological estimate of that property is determined (*e.g.*, in ICES it is Blim – damaged productivity). The next issue is estimate the current status relative to that point with some measure of uncertainty, so a buffer is needed (*e.g.*, Bpa). This is the point where the probability that true stock biomass may be at the limit exceeds 0.05. This framework allows the current value of an indicator to guide risk-averse management, and makes the whole system precautionary.

The issue of predictability requires us to consider various temporal scales of interest. It is also instructive to consider whether "scenario explorations" associated with climate change and with marine ecosystem dynamics have important differences. Climate change has no expectation of accuracy on timescales greater than 30 days or less than 30 years. Ecosystem dynamics, on the other hand, at lead times of 3 to 7 (10?) years provide some of the core decision support for management. The climate change decisions are long-term strategic, but the ecosystem dynamics decisions are medium-term and tactical.

In considering what to predict, it seems that one should not try to capture inter-annual flutter. It will be more important to know how the probability of an extreme event (good or bad) varies with natural or anthropogenic forcers, rather than try to predict minor deviations up or down from long-term average conditions Multi-factor non-parametric probability density estimation methods do show inflections in plot of P (extreme event) as f (specified forcers). The predictions should be easy to use and to interpret but they do require decisions about what is "extreme."

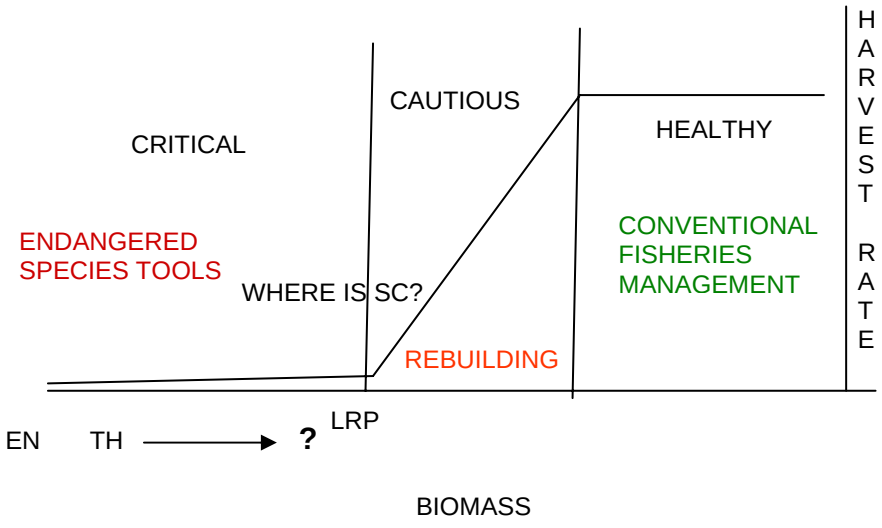

Fig. 18 The classic 3-stage model.

Ecosystem indicators have also been a priority for OSPAR (the Oslo–Paris Commission) to use in fulfilling their mandate for protection of environmental quality of the North East Atlantic. ICES was requested to advise on the suitability of different sorts of ecosystem indicators, and the Working Group on Ecosystem Effects of Fishing was asked to undertake the evaluation. Over a series of several meetings in the late 1990s and early 2000s, they developed screening criteria for ecosystem indicators, reviewed literature on marine ecosystem indicators, and tested both their criteria and a number of classes of indicators with some extensive data sets from the ICES area. Starting with a suite of more than 60 types of indicators, the ICES Working Group on Ecosystems (WGECO) found that the best alternatives included:

For the biodiversity/fish community

- slope of size-spectrum;
- mean length of fishes from a standardized survey;
- % of fish greater than some system-specific size in a standardized survey;
- bycatch rate of "particularly sensitive" species in observer data, where "particularly sensitive" is determined by rough estimates of "q" for the gear and an estimate of sustainable Z from life history parameters and
- survey-based abundance estimates;
- K-dominance (ABC) curves;
- frequency distribution of L_{max} in a standardized survey;
- species richness.

For trophodynamic processes/status

- No model-based indicators were found to perform well, and size-based indicators are better, even though they are surrogates for the processes.

For spatial integrity

- No suitable indicators were found.

Group discussion

James Overland: I have a hard time imagining applying the psychometric analogue to marine ecosystems. Where might we find a significant population of marine ecosystems? Do we need to compute a pdf of ecosystem responses?

Jake Rice: No one has tried to do this…it works in psychometrics.

Jason Link: The leads like a commercial for the comparative ecosystem session at the 2007 ICES Annual Science Conference that will try to pull all of the high latitude ecosystem comparisons together (Convenors: Ian Perry, Bernard Megrey, Jason Link).

Andrea Belgrano: How do they deal with the multi-dimensional issues that are so critical to the study of ecosystems, in psychometrics?

Rice: They would argue that human personality is a dimensionally complex problem.

Overland: I think that it relates to overfishing; our problem is shifted to looking at the shift in the system dynamics or response to climate, *etc.* which may eventually have a management implication but does not have one right now. Is there a different way that we should be thinking about things?

Rice: Indicators are used widely in environmental health reporting, *e.g.*, coastal pollution. There they ask, "What is the optimal way to use a community that has this common trait compared to one that does not?"

Francis Wiese: I like the idea of retrospective studies of indicators.

Rice: For regimes, you do not want to base decision making on insensitive indicators. The indicator needs to have a history.