

The state of the art in stock assessment: where we are and where we are going*

RAY HILBORN

School of Aquatic and Fisheries Sciences, Box 355020, University of Washington, Seattle, WA 98195. USA.
E-mail rayh@u.washington.edu

SUMMARY: Throughout the world's major commercial fisheries the standard paradigm includes fitting a stock assessment model to the data and then applying some form of reference point or a range of reference points, usually a target exploitation rate, to calculate a point estimate of the annual allowable catch. Over the last decade the methods available to be used in these models have changed dramatically from those using only catch, catch-at-age and survey or CPUE data to methods that now use every source of data available in a totally integrated framework. The use of meta-analysis has provided formal statistical methods for incorporating information learned from other stocks. Modern methods make it possible to express uncertainty in all model outputs, but management agencies are only now learning to deal with explicit recognition of uncertainty and have lagged far behind the scientific capability to express uncertainty. While modern stock assessment models have grown increasingly complex and their development is limited to a priesthood of experts, I believe the future trend will be to base management decisions on simple rules that are more often data-based rather than model-based, while the complex models will serve primarily to evaluate the robustness of these decision rules.

Key words: stock assessment, fisheries management, fisheries models.

INTRODUCTION

The purpose of this paper is to review the recent trends in stock assessment and to look at where it is going now and where it ought to be going in the future. This is not a comprehensive review of stock assessment methods but a personal interpretation of the present and where I think we should be headed. Recognising that stock assessment in Europe and Eastern North America has evolved somewhat differently from that in the Pacific where most of my experience has been, some of my comments may seem regional.

*Received December 6, 2000. Accepted November 27, 2001.

The purpose of stock assessment is to provide support for decision making by (1) describing alternative possible states of nature, (2) determining the consequences of taking different management actions under different states of nature, and (3) calculating the probability of different states of nature. The evolution of modern stock assessment has been directed towards providing the best possible technical support to tasks 1-3. In the sections below I outline what I see to be the five key characteristics of modern stock assessments. Generally these are new developments within the last 10-20 years, and the newer ones are far from universally practiced –indeed I doubt that there are more than a handful of assessments around the world that use all these approaches.

THE KEY CHARACTERISTICS OF MODERN STOCK ASSESSMENT

Integration of all information

The classic stock-assessment methods of the 1960s and 1970s were very restricted in the data used: VPA and statistical catch-at-age methods used only catch-at-age data and some tuning index such as surveys or CPUE; spawner recruit analysis used only spawner recruit data; and biomass dynamics models used CPUE and catch (Hilborn and Walters 1992). Modern methods use all available information in a unified framework and may simultaneously include surveys, CPUE, age-distributions, length distributions, and tagging (McAllister *et al.*, 1994, Punt and Kennedy, 1997). Methot's stock synthesis model (Methot, 1989) was the pioneer in such models, but now fitting to all available data has become commonplace as scientists seek to use the models to capture all knowledge about stock size and productivity.

Expressing uncertainty

In the past the dominant product of a stock assessment was a point estimate of a quantity of interest, usually population size, maximum sustainable yield, or potential yield based on a target exploitation rate. It is becoming increasingly common to provide measures of uncertainty in the outputs of stock assessments, with the three dominant approaches being (1) sensitivity analysis, (2) bootstrapping to produce distributions, and (3) Bayesian posterior distributions. All three methods are popular at present although bootstrapping is fading and Bayesian methods are becoming more popular.

In sensitivity analysis different assumptions regarding parameter values or inclusion of different data sets are run and results are compared. This is standard practice even when bootstrapping or Bayesian methods are used to calculate distributions. However useful sensitivity analysis may be for an analyst to understand the behaviour of the model, sensitivity analysis poses serious problems in relaying management advice if the results are sensitive to assumptions. For example, let us suppose that the estimated stock size or MSY depends on which data sets are used, and this uncertainty is relayed to decision makers. How do they interpret the uncertainty? The fear is, of course, that they will simply "choose" the case they like the most. Sensitivity analysis, in

the absence of scientifically based probabilities on the different assumptions, can only serve to confuse decision makers. My suggestion is that if results are sensitive to assumptions, the analysts must assign a probability to the alternative assumptions using the best available scientific understanding.

Bootstrapping and Bayesian methods are both used to obtain distributions of outputs of interest that can then be used as inputs to formal decision methods. The key is that they both attempt to assign relative probabilities to possible states of nature—one of the objectives of stock assessment. Bootstrapping has the advantage that it is computationally straightforward, but the disadvantage is that there is no theoretical basis for using the results as probabilities. Statisticians acknowledge that the outputs of bootstrapping should not be interpreted as probabilities, yet many stock assessment authors use them as such out of convenience.

Bayesian methods do produce statistically rigorous probabilities but have the disadvantages of computational complexity and technical problems in defining proper prior distributions. Major progress has been made in computation in the last 10 years, and there has been equal progress in formulating data-based rather than subjective priors. It is for this reason as well as their simple intuitive explanation as probabilities, I believe, that Bayesian methods are growing in popularity (McAllister *et al.*, 1994, Punt and Hilborn, 1997).

In many assessments the analysts simply stop at presenting uncertainty and go no further. But increasingly the results are being taken into formal decision analysis where alternative states of nature are explicitly considered and the consequences of different actions under different states of nature are presented to decision makers.

The present: management by reference points

If any "standard" practice in linking stock assessment to decision making has evolved, it is taking the best estimated stock size from an assessment and calculating a recommended harvest by multiplying this stock size by a desired exploitation rate that may change in relation to the stock size (Restrepo and Powers, 1999). In both the U.S. and Canada a paradigm has evolved of reference points based on current stock size in relation to a hypothetical unexploited stock size as shown in Figure 1. This is a slight simplification of the rule adopted by the Pacific Fisheries Management Council in the western

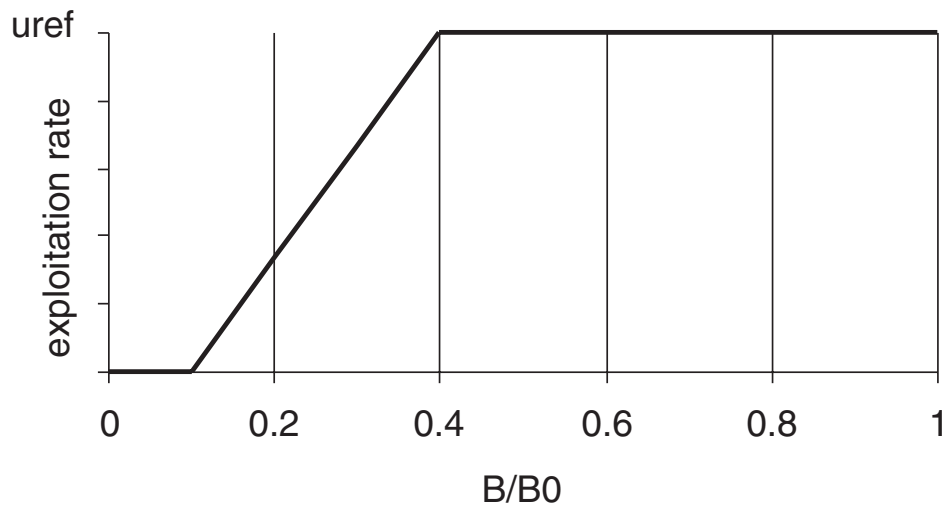


FIG. 1. – The relationship between stock size and target exploitation rate, slightly adapted from the rule used by the Pacific Fisheries Management Council. Uref is the target exploitation rate, B is the current estimated stock size and B0 is the estimated unfished equilibrium stock size. In some applications exploitation rates are instantaneous fishing mortality rates, in other applications they are catch divided by biomass.

states of the U.S. For each species a target exploitation rate (Uref) and a virgin biomass are defined, and then the recommended TAC is taken by multiplying the best estimated stock size by the target exploitation rate. When uncertainty in current stock size is explicitly considered, one can calculate the best recommended TAC by weighting the probabilities the assessment assigns to different stock sizes. Reference points may be calculated for exploitation rates, or stock biomass. It is becoming increasingly common to use or at least consider precautionary reference points, whose general characteristic is that they aim for larger stock biomass and lower exploitation rates than reference points based on traditional maximum yield objectives (Hilborn *et al.*, 2001).

There are several key problems with management by reference points, particularly the often large uncertainty in actual stock size, and even larger uncertainty in unfished biomass. While some (Hilborn, 2002) have argued we need to move away from reference points, they are, at present, a common feature of assessments and management.

Including environmental change

In the past we assumed production parameters were time-invariant, but increasingly we recognise and incorporate changes over time in the production parameters of our models. Cushing's (1982) "Climate and fisheries" was probably the seminal work in this change, while the clear climate impact in the California sardine (Jacobson and MacCall, 1995) and the impact of the Pacific Decadal Oscillation

greatly influenced work in the Pacific (Hare and Francis, 1995). Similar clear environmentally induced changes in abundance in other parts of the world, particularly Europe, have led to much more consideration of time-varying parameters. In western North America we tend to consider environmental change as regimes, but as yet few of the assessments make explicit allowance for such change. When they do, the assessments allow for different average recruitments during different regimes.

In Europe, there has been more of a tendency to include environmental covariates as part of the assessment models, either as prey or predator impacts on natural mortality rates as found in multi-species VPA, or using a physical parameter such as temperature in recruitment relationships. Myers (1998) did a review of temperature recruitment correlations and found that they tended to hold up over long periods of time only at the northern and southern ends of the distribution of the species.

When we include the possibility of environmental change in an assessment, we open up further doors of uncertainty. We are obviously going to be less certain of the consequences of different management actions if the future environment is uncertain. For instance, if a decline in recruitment has occurred at the same time as a decline in stock size and an environmental change, the scientists would be less confident that rebuilding stock size would increase recruitment. If we accept an environmental explanation for change in recruitment, then the appropriate policy might simply be to maintain stock size and wait for the environment to change.

Statistical time-series methods

A particularly elegant method for dealing with parameters that change over time is the incorporation of statistical time series methods into stock assessment models by Fournier and Archibald (1982), expanded on by Haist *et al.* (1994) and others (e.g. Fournier and Hampton, 1996). In statistical time-series analysis, parameters such as catchabilities, age-specific vulnerabilities, and spawner-recruit parameters are allowed to change over time, but are subject to constraints. For instance, in the traditional analysis, the scaling factor between CPUE and abundance is governed by a relations $I=Uq$ where I is the index, U is the CPUE and q is the scalar. The parameter q is usually assumed to be constant over time. It is possible, of course, to allow q to be a free parameter in every year of the model, but this would mean that the index provides no information about abundance. In statistical time-series analysis, the changes in q are constrained such that

$$\log(q_{t+1}) = \log(q_t) + \Delta_{qt}$$

where Δ_{qt} is the amount the q changes from time to time and is constrained by the relationship

$$\Delta_{qt} \approx N(0, \sigma_{\Delta}).$$

The model is thus that q_t is a random walk with normally distributed steps. If all Δ_{qt} are zero, there is no additional “penalty” in the likelihood. Similarly, if we set σ_{Δ} to be very large, there are no penalties and the q ’s can change from year to year with no constraint, but if we set σ_{Δ} small, the q ’s cannot change much from year to year.

Statistical time series methods allow the same model to mimic VPA methods where age and year specific fishing mortality rates are free parameters (as in ADAPT (Gavaris, 1988)) by letting age-specific selectivity change freely from year to year, and catch-at-age methods such as CAGEAN (Deriso *et al.*, 1985) and derivatives in which age-specific selectivity is assumed constant, and to represent intermediate assumptions about changes in selectivity over time.

Statistical time-series methods are not in wide use—indeed they pose significant computational problems because every Δ_{qt} is a parameter to be estimated, and if many of the standard parameters of the model are allowed to change over time, many hundreds of parameters need to be estimated. In the U.S.

National Research Council report (NRC 1998) statistical time-series methods were used to try to detect changing parameters over time and they appear to hold considerable promise.

Meta analysis

The traditional approach has been to assume that many parameters of our models were fixed and known without error, particularly natural mortality rates and spawner-recruit parameters. This was not because we really believed that we knew these parameters, but rather that the data in our assessments did not provide much information about them. The approach of meta-analysis (Hilborn and Liermann, 1998), using a large number of data sets to define distributions of parameters, has been growing in popularity. Pauly’s well known work on natural mortality rates (Pauly, 1980) was the first meta-analysis that became popular. Myers (2000) and Myers *et al.* (2001) have performed meta-analysis of spawner recruit parameters, and Liermann and Hilborn (1997) of depensation. Other recent examples include Dorn’s (2001) analysis of spawner-recruit parameters for rockfish on the pacific coast of the U.S. and the analysis of intensity of compensatory mortality to evaluate power plant impacts in the Delaware River by Myers *et al.* (2001).

The product of a meta-analysis is a distribution of the parameter that can be used either as a prior probability distribution in a Bayesian setting or as an additional likelihood component in a maximum likelihood model.

There is a great opportunity at present for meta-analysis of other important parameters including spawner-recruit variability, gear efficiency in trawl surveys, rates of change in parameters due to environmental change and undoubtedly many others.

THE FUTURE

The growth in our computational power and modelling sophistication allows us to do estimations we couldn’t dream of 10 years ago, and we can look forward to large Bayesian models incorporating meta-analytic results that will allow for all sorts of environmental change and be exceedingly powerful and general. This will undoubtedly happen, but I believe the models will become increasingly less central to the regulation-setting process.

The growing complexity has a number of negative features. First is the lack of transparency: increasing complexity and importance of internal assumptions make it often hard to understand what drives many assessments. As assessments become more complex the models are no longer simplifications to enable understanding but become complex black boxes. When decision makers ask “why does the model say this” the analyst is more likely not to have a simple answer. Second is lack of access: fewer and fewer people are able to “play the game” at the highest level. In the US, for instance, the National Marine Fisheries Service has declared a national manpower shortage in quantitative population dynamics. There simply are not enough trained people in the US to do the modelling and analysis at the level that is expected. More assessments are being performed by people who really do not understand the details of the computer software they are using. More importantly, as the models become more complex, with fewer people able to run and understand them, there will be less understanding of the models within organisations and a tendency to resort to analytic methods or rules of thumb that are understandable to a wider group of participants.

However, the BIG problem is that as modelling has become so central to many decision making processes we have lost sight of what is truly important: the data that go into the assessments (Rose, 1997). In many fisheries I work in, most of the energy goes into modelling and analysis and a few ongoing data collection programs. Few scientists are working on biology and new data; and few scientists are in touch with the fishery and what is actually happening on the water. The institutional and legal requirements for stock assessment, particularly in the U.S., means that our actual contact with and understanding of fisheries is diminishing rapidly, and we are playing technical games with models that are becoming less relevant to the real fishery.

Thus, I foresee the end of stock assessment as we know it, the end to running models each year to produce an estimate of stock size (or a distribution) that is then used to determine management actions. I believe we should and will move towards using management procedures (Butterworth and Punt 1999) in which regulations are modified using rules that directly use data or very simple models. Highly complex models will be relegated to the role of providing alternative states of nature and their probabilities and will be used to test the management procedures for robustness.

For those unfamiliar with management procedures, they are a set of rules that specify (1) data to be collected, (2) how data will be processed including simple models and (3) how decisions will change in relation to the data. Many fisheries are managed by setting the TAC as a harvest rate times the stock size as estimated in a stock assessment. This resembles a management procedure, except that normally the assumptions of the stock assessment are free to be adjusted by scientists each year. On the other hand, in a management procedure the equations and assumptions and data inputs are specified ahead of time.

The benefits of management procedures are primarily that they are transparent and thus acceptable to user groups, and can be defined to allow for both conservation of the biological resource and social and economic returns from the fishery.

CONCLUSION

Throughout the fisheries management institutions of the world, stock assessment models have become central to fisheries decision making, and these models are increasingly encompassing a broader range of data and admitting more uncertainty. I see a major trend away from these models being the centerpiece of harvest regulation and a move towards simpler rules for setting harvest levels, with the complex models being used primarily to test the robustness of the rules. I would like to see—but do not see much evidence of it at present—more effort devoted to biological understanding of the resources and better understanding of the dynamics of the fishery by those involved in fisheries management.

REFERENCES

- Butterworth, D.S. and A.E. Punt. – 1999. Experience in the evaluation and implementation of management procedures. *ICES J. Mar. Sci.*, 56: 985-998.
- Cushing, D. – 1982. *Climate and fisheries*. Academic Press, London.
- Deriso, R.B., T.J. Quinn II and P.R. Neal. – 1985. Catch-age analysis with auxiliary information. *Can. J. Fish. Aquat. Sci.*, 42: 815-824.
- Dorn, M. – (In press). Advice on harvest rates for west coast rockfish based on Bayesian meta analysis. *N Am. J. Fish. Mgmt.*
- Fournier, D.A. and C.P. Archibald. – 1982. A general theory for analyzing catch-at-age data. *Can. J. Fish. Aquat. Sci.*, 39: 1195-1207.
- Fournier, D.A. and J. Hampton. – 1996. Preliminary analyses of South Pacific albacore catch, effort and length-frequency data, using an age-structured, length-based model with spatial structure. Paper presented to ICCAT 25th anniversary tuna symposium, June 10-18 1996, Azores, Portugal. Also available at <http://www.island.net/~otter>.

- Fournier, D.A., J.R. Sibert, J.R. Majkowski and J. Hampton. – 1990. MULTIFAN: a likelihood-based method for estimating growth parameters and age composition from multiple length frequency data sets illustrated using data for southern bluefin tuna (*Thunnus maccoyii*). *Can. J. Fish. Aquat. Sci.*, 47: 301-317.
- Gavaris, S. – 1988. An adaptive framework for the estimation of population size. *Can. Atl. Fish. Sci. Adv. Comm. Res., Document* 88/29. 12 pp.
- Haist, V., D. Fournier and J. Schweigart. – 1993. Estimation of density-dependent natural mortality in British Columbia herring stocks through SSPA and its impact on sustainable harvesting strategies. In S.J. Smith, J.J. Hunt, and D. Rivard (Eds.), *Risk evaluation and biological reference points for fisheries management. Can. Spec. Publ. Fish. Aquat. Sci.*, 120.
- Hare, S.R., and R.C. Francis. – 1995. Climate change and salmon production in the Northeast Pacific Ocean. *Can. Spec. Pub. Fish. Aquat. Sci.*, 121: 357-372.
- Hilborn, R. – 2002. The dark side of reference points. *Bull. Mar. Sci.*, 10: 403-408.
- Hilborn, R. and M. Liermann. – 1998. Standing on the shoulders of giants: learning from experience. *Rev. Fish Biol. Fisheries*, 8: 273-283.
- Hilborn, R., J.J. Maguire, A.M. Parma, and A.A. Rosenberg. – 2001. The precautionary approach and risk management: can they increase the probability of success in fisheries management. *Can. J. Fish. Aquat. Sci.*, 58: 99-107.
- Hilborn R. and C.J. Walters. – 1992. *Quantitative fisheries stock assessment and management: choice, dynamics and uncertainty*. Chapman and Hall, New York 570 pps.
- Jacobson, L.D. and A.D. MacCall. – 1995. Stock-recruitment models for Pacific sardine (*Sardinops sagax*). *Can. J. Fish. Aquat. Sci.*, 52: 566-577.
- Liermann, M. and R. Hilborn. – 1997. Depensation in fish stocks: a hierarchic Bayesian meta-analysis. *Can. J. Fish. Aquat. Sci.*, 54: 1976-1984.
- McAllister, M.K., E.K. Pikitch, A.E. Punt and R. Hilborn. – 1994. A Bayesian approach to stock assessment and harvest decisions using the sampling/importance resampling algorithm. *Can. J. Fish. Aquat. Sci.*, 51: 2673-2687.
- Method, R.D. – 1989 Synthetic estimates of historical abundance and mortality for northern anchovy. *Am. Fish. Soc. Symp.*, 6: 66-82.
- Myers, R.A. – 1998. When do environment-recruit correlations work? *Rev. Fish. Biol. Fisher.*, 8: 285-305.
- Myers, R.A. – (In press). So what do we really know about recruitment? Generalizations about maximum reproductive rate, density dependence and variability. *ICES J. Mar. Sci.*
- Myers, R.A., N.J. Barrowman, R. Hilborn and D.G. Kehler. (In press). – Inferring the Bayes priors with limited direct data with applications for risk analysis and reference points. *N. Am. J. Fish. Manage.*
- National Research Council. – 1998. *Improving fish stock assessment*. NRC Press, Washington D.C.
- Pauly, D. – 1980. On the interrelationships between natural mortality, growth parameters, and mean environmental temperature in 175 fish stocks. *J. Cons. int. Explor Mer.*, 39: 175-192.
- Punt, A.E. and R. Hilborn. – 1997. Fisheries stock assessment and decision analysis: the Bayesian approach. *Rev. Fish Biol. Fisheries*, 7: 35-63.
- Punt, A.E. and R.B. Kennedy. – 1997. Population modelling of Tasmanian rock lobster, *Jasus edwardsii*, resources. *Mar. Freshwater Res.*, 48: 967-980.
- Restrepo, V.R. and J.E. Powers. – 1999. Precautionary control rules in US fisheries management: specification and performance. *ICES J. Mar. Sci.*, 56: 846-852.
- Rose, G.A. – 1997. The trouble with fisheries science. *Rev. Fish Biol. Fisheries*, 7: 365-370.