Simple representation of the dynamics of biomass error propagation for stock assessment models

Carl Walters

Abstract: Errors in estimates of current exploitable biomass from age-structured stock assessment models typically propagate over time as though they had been computed by using a very simple Kalman filter equation. This equation shows that persistent overestimates of biomass during stock declines are caused mainly by three factors: first-order autoregressive effects of past errors, hyperstability in relative abundance indices, and overestimates of current recruitment. Besides exposing these main causes of bias in biomass estimation, the equation can be used to develop very efficient closed-loop or management performance simulations and to provide timely biomass estimates in situations where it is not practical or economical to carry out full assessments every year.

Résumé : Les erreurs des estimations de biomasse courante exploitable dans les modèles d'évaluation de stock avec structure en âge se répercutent généralement dans le temps comme si elles avaient été calculées à l'aide d'une équation très simple du filtre de Kalman. Cette équation démontre que des surestimations répétées de la biomasse alors que les stocks diminuent peuvent être dues à trois facteurs, soit, les effets d'autorégression de premier ordre des erreurs du passé, l'hyperstabilité des indices d'abondance relative et la surestimation du recrutement actuel. En plus de mettre en lumière ces causes principales d'erreur dans les estimations de biomasse, l'équation peut servir à mettre au point des simulations très efficaces en boucle fermée ou de performance de gestion et elle peut fournir des estimations de la biomasse en temps opportun lorsqu'il n'est pas pratique ou économique de faire une évaluation complète chaque année.

[Traduit par la Rédaction]

Introduction

Modern stock assessments are usually presented with measures of uncertainty for key parameters, such as probability distributions for current stock size. Such measures can obviously be helpful in design of precautionary harvesting policies. But it is not always made clear that estimation errors are not independent of one another over time (because successive assessments use the same data, along with incremental new information), leading to the possibility of persistent periods of over- or under-harvesting when the assessment results are applied over time. For example, persistent overestimation of stock size apparently contributed to the demise of the northern cod (*Gadus morhua*) stock (Walters and McGuire 1996). Impacts of such persistent errors only become apparent when assessment is accompanied by closed-loop (Walters 1986) or "management procedure" (Butterworth et al. 1997; Cooke 1999; Punt and Smith 1999) simulations of future data gathering, assessment, and policy implementation.

C. Walters. Fisheries Centre, University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: c.walters@fisheries.ubc.ca).

This paper shows that persistence of errors in the estimation of current biomass, a critical issue for stocks that are managed by output controls (total allowable catch (TAC) or quota regulation), can be predicted even for complex stock assessment methods by using a very simple model based on the theory of linear or Kalman filtering (Gelb 1974; Pella 1993; Reed and Simons 1996). This model can be used to construct efficient management procedure simulations and to demonstrate the main causes of dangerous upward bias in biomass estimates so as to help justify improvements in monitoring and prediction methods.

Linear filter approximations for updating stock size estimates

For estimation of a single scalar quantity, vulnerable biomass B_t, using a single abundance index y_t, where y_t is assumed proportional on average to B_t ($y_t = qB_t$), the elaborate theory of state estimation for linear dynamical systems reduces to the simple Kalman filter equation

$$(1) \qquad B_{t|t} = B_{t|t-1} + K_t(y_t - q_{t-1}B_{t|t-1})$$

$$= B_{t|t-1} + K_t^*(B_t^* - B_{t|t-1})$$

Here, $B_{t|t}$ is the best (minimum variance) estimate of B_t given all of the data to time t, $B_{t|t-1}$ is the best estimate or prediction of B_t given only the data through time $t-1$, q_{t-1} is

the best estimate of q given data to time $t - 1$, K_t is the filter or Kalman gain. The second (and simpler) form is to use y_t to provide a current estimate $B_t^* = y_t / q_{t-1}$ based only on the year t relative abundance data, then update this using the alternative filter gain $K_t^* = q_{t-1} K_t$. In both cases, the basic idea is to first predict B_t using past data ($B_{t|t-1}$) and then correct this prediction in proportion to its error in predicting the data ($y_t - q_{t-1} B_{t|t-1}$ or $B_t^* - B_{t|t-1}$). Because B_t^* and $B_{t|t-1}$ are independent (except through weak effects of estimation of q_{t-1}), the variance of the current estimate $B_{t|t}$ is

$$(2) \qquad \sigma^2_{B_{t|t}} = (1 - q_{t-1} K_t)^2 \, \sigma^2_{B_{t|t-1}} + K_t^2 \, \sigma^2_y$$
$$= (1 - K_t^*)^2 \, \sigma^2_{B_{t|t-1}} + K_t^{*2} \, \sigma^2_{B_t^*}$$

Differentiating this variance with respect to K_t or K_t^*, it is easily seen that variance of $B_{t|t}$ is minimized by setting K_t equal to (i.e., balancing the use of the past and current data by setting K_t or K_t^* to)

$$(3) \qquad K_t = \frac{q_{t-1} \sigma^2_{B_{t|t-1}}}{q_{t-1} \sigma^2_{B_{t|t-1}} + \sigma^2_y / q_{t-1}}, \qquad K^* = \frac{\sigma^2_{B_{t|t-1}}}{\sigma^2_{B_{t|t-1}} + \sigma^2_{B_t^*}}$$

which is the familiar expression for the Kalman gain in scalar cases ($\sigma^2_{B^*} = \sigma^2_y / q_{t-1}^2$ and $\sigma^2_{B_{t|t-1}}$ is the variance of biomass prediction at time t given data up to time $t - 1$). A very large literature on the theory of optimal stochastic control shows how eqs. 1 and 3 extend to multivariate cases and why the Kalman filter provides optimal state estimation performance for a wide variety of problems where the system dynamics are linear ($B_{t|t-1}$ is a linear function of $B_{t-1|t-1}$ with known parameters, known matrix analog of q for multiple observations with known covariance matrix) and the control objective can be represented by a quadratic loss function (variance).

No particular assumption is made in eq. 1 about how the biomass predictions $B_{t|t-1}$ are made. Using estimated numbers at age $N_{a,t|t}$ (estimated numbers of age a animals at year t given data up to time t) from age-structured stock assessments and assuming numbers at age can be approximated by the removal–survival model $N_{a+1,t+1} = S(N_{a,t} - C_{a,t})$ with constant survival $S = e^{-M}$, we would use a prediction of the form

$$(4) \qquad B_{t|t-1} = w_1 v_1 N_{1t|t-1} + \sum_{a=2}^{a_{max}} w_a v_a S_a (N_{a-1,t-1|t-1} - C_{a-1,t-1})$$

which includes predictions of survivors from each age along with a prediction of new recruits and requires estimated age schedules of body weights w_a and vulnerabilities v_a. An important point that will be used later to develop simple models for estimation error propagation is that that the predictions of eq. 4 can most often be closely approximated by the delay-difference result

$$(5) \qquad B_{t|t-1} = g_{t-1}(B_{t-1|t-1} - H_{t-1}) + w_k N_{kt|t-1}$$

where H_{t-1} is the total biomass harvest $\Sigma_a w_a C_{a-1,t-1}$, k is the age at which $v_a = 0.5$, and g_t is the growth–survival constant

$$(6) \qquad g_t = S\left(\frac{\alpha}{\overline{w}_t} + \rho\right)$$

Here, α and ρ are the slope and intercept of the Ford-Brody growth model $w_{a+1} = \alpha + \rho w_a$ estimated using weights for age k and older fish, $S = e^{-M}$ is the mean survival rate of age k and older fish, and $\overline{w}_t$ is the mean weight of age k and older fish in the year t catch. Equation 5 is derived by replacing the vulnerabilities v_a with the knife-edged assumption ($v_a = 0$ for $a < k$, $v_a = 1$ afterwards), but it generally works very well for cases where vulnerability increases in a logistic pattern with age (it can fail badly when vulnerability is dome-shaped). In most cases, errors in forecasts ($B_{t|t-1}$) due to the g_t approximation will be modest compared with errors caused by inaccurate or biased predictions of recruitments $N_{kt|t-1}$.

The Kalman gains predicted by eq. 3 are typically large for low values of t (B_t is highly uncertain early in fishery development), but rapidly stabilize at nearly constant values associated with persistent uncertainty in the prediction of $B_{t|t-1}$. That persistent uncertainty arises partly from observation errors (in y_t, age-specific catches $C_{a,t}$, and changes in vulnerability v_a) and partly from notorious difficulty in predicting recruitments $N_{1,t}$, $N_{k,t}$. Using the approximation eq. 5, it is apparent that $\sigma^2_{B_{t|t-1}}$ will vary over time as

$$(7) \qquad \sigma^2_{B_{t|t-1}} = g_{t-1}^2 \, \sigma^2_{B_{t-1|t-1}} + \sigma^2_{N_{k,t}}$$

for which the dominant term (at least for large t) is typically the recruitment variance $\sigma^2_{N_{1,t}}$. However, it should be noted that K_t^* does not predictably increase as the variance of y_t decreases because $\sigma^2_{B_{t|t-1}}$ also decreases with reductions in survey variance.

Considering the complexity of population dynamics assumptions and nonlinear estimation methods used in age-structured stock assessment procedures, there is no obvious reason to expect the results from such procedures to behave over time in the simple way predicted by eq. 1. But there is a very simple way to test whether a given procedure does in fact exhibit simple behavior at least for times t large enough that the Kalman gain K_t^* becomes approximately time-independent. For constant K^*, eq. 1 can be rearranged in the regression format

$$(8) \qquad B_{t|t} - B_{t|t-1} = K^*(B_t^* - B_{t|t-1})$$

If the estimates $B_{t|t}$, $B_{t|t-1}$, and B_t^* from the assessment procedure show this simple proportional relationship when the procedure (and its eq. 5 forecasts) is run with increasing numbers of years t of data, then the procedure does behave as though it were a simple filtering operation, and regression slope estimate of K^* can be used to construct a simple simulation (using eq. 1) of the procedure. An alternate, even simpler test is to calculate the apparent filter gains $K_t^* = (B_{t|t} - B_{t|t-1})/(B_t^* - B_{t|t-1})$ using the B estimates from the procedure and then check to see if these estimates converge over time to a single, stable value.

Sample results of the eq. 8 regression are shown (Fig. 1) from application of virtual population analysis (VPA) and stock synthesis (SCA) age-structured assessment procedures to simulated catch-at-age and survey data for a population with $S = 0.8$, increasing exploitation rate over time (from 0.05 to 0.3 over 20 years), and highly variable recruitment (lognormal with standard deviation 0.6). For this example,

Fig. 1. Typical simulated patterns of change in vulnerable biomass estimates and estimation errors for two age-structured stock assessment methods (VPA, virtual population analysis; SCA, statistical catch-at-age or stock synthesis). Results are shown for years 5 to 20 from 10 simulation trials. Top panels demonstrate that updated biomass estimates $B_{t|t}$ can be closely predicted by Kalman filter equations using prediction $B_{t|t-1}$ from previous data and current estimate B_t^* based on survey data for year t; slopes of regressions are estimates of filter gain K^*. Bottom panel shows updated estimates $B_{t|t}$ versus "true" biomasses B_{true}.
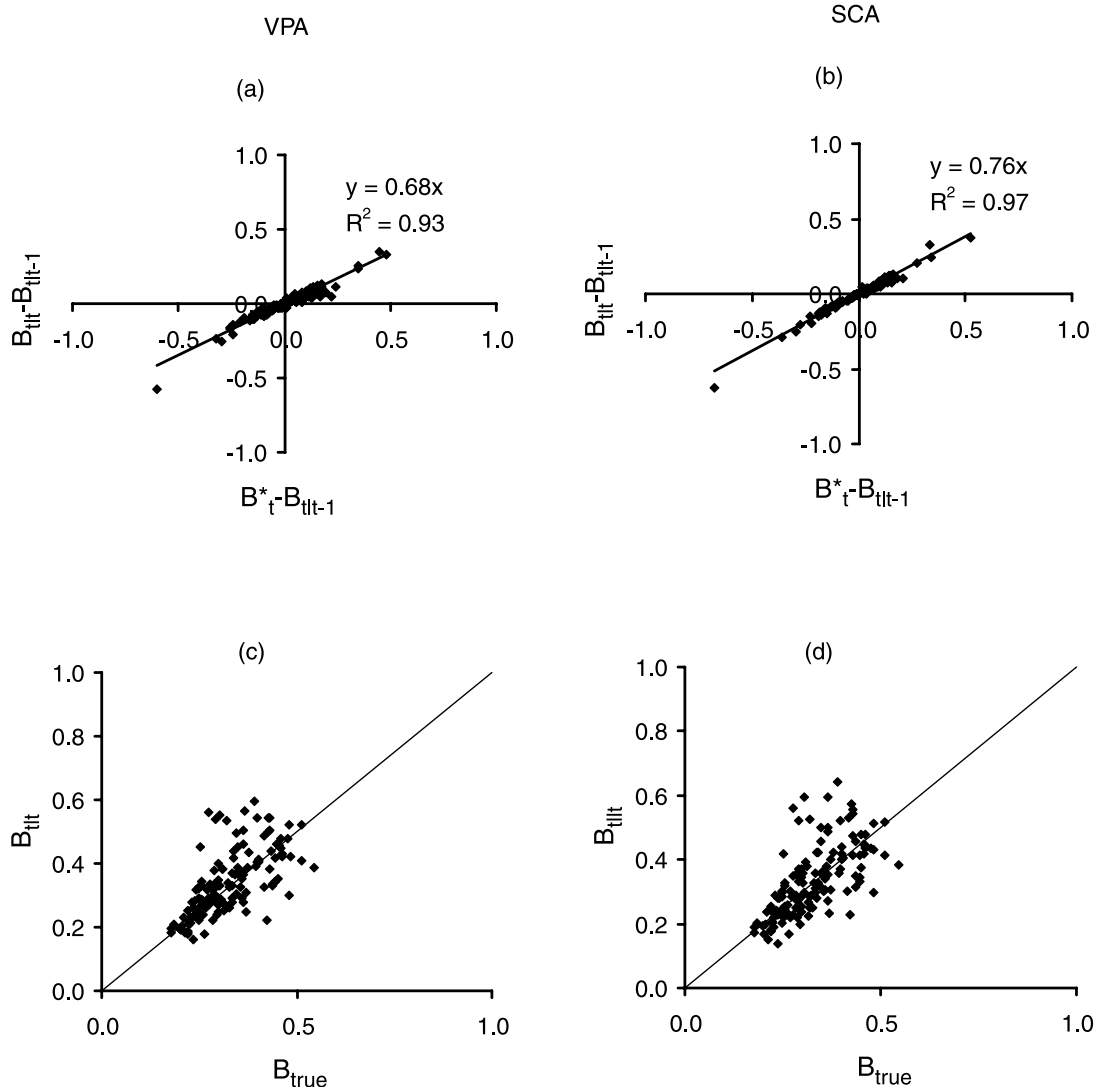
simulated catches at age were obtained by taking multinomial age composition samples of 800 fish per year (aging method assumed accurate), and annual biomass survey data were generated with lognormal errors and a coefficient of variation (CV) of 0.3. The VPA analyses were done with the noniterated method for estimating terminal vulnerabilities $v_{a,\text{T}}$ recommended by Walters and Punt (1994), which involves "tuning" to survey data but avoids some problems with programs like ADAPT in estimation of terminal vulnerabilities. The SCA was done by fitting both catch-at-age and survey data using square-root transformations of the catch-at-age data and lognormal error assumptions for survey data, as recommended by Quinn and Deriso (1999). Vulnerabilities at age were simulated to be, and were assumed in the SCA to be, logistic with age and stable over time. Similar results were obtained with more complex examples where vulnerabilities changed over time so as to subject younger

fish to more intense exploitation as the simulated stock declined.

Results very similar to those of Fig. 1 were obtained for simulation trials involving a wide range of temporal harvest rate patterns, including cases where vulnerability schedules were deliberately varied over time (e.g., to make younger fish more vulnerable to harvesting during stock declines) and (or) simulated survey observations were generated from a hyperstable function ($y_t = q B_t^\beta$, $\beta \ll 1.0$) so as to deliberately cause upward bias in the $B_{t|t}$ estimates. In every case, the linear filter eq. 1 proved to be an excellent "predictor" of the effect of new relative abundance observations y_t (or B_t^*) on the updated stock size estimates $B_{t|t}$, and the regression estimates of K^* from eq. 8 were typically in the range 0.5–0.7 (unless unrealistically low CVs were assumed for y_t). Variation around the eq. 8 prediction was typically a bit higher for VPA than for SCA, apparently due to using $N_{a,t|t}$

calculated directly from variable catch estimates $C_{a,t}$ ($N_{a,t|t} = C_{a,t}/v_a u_t$) in the VPA calculations of $B_{t|t-1}$ using eq. 4. For both methods, results for $t < 5$ were also more erratic, most likely due to variation in the conditional maximum likelihood estimates q_{t-1}. Average estimation errors (as measured by differences between $B_{t|t}$ and the true B_t) were similar for both methods (Fig. 1). It did not seem to matter to the performance of eqs. 1 or 8 as predictors of $B_{t|t}$ that the simulated variance of y_t was proportional to B_t rather than constant as assumed in derivations like eq. 3 for optimum filter gains.

One practical implication of these findings is that the complex nonlinear estimation procedures used in SCA (with n ages over t years, the SCA search is for at least $2t + n - 1 + n/2$ exploitation rates, initial $N_{a,t}$, and v_a parameters) may commonly be a waste of time compared with the one-parameter (u_t) search needed for VPA. In other words, it is misleading to represent the high parameter counts of SCA as somehow a virtue or predictor of more realistic results from that approach. VPA "failures" like those reported for ADAPT in NRC (1998) are most likely due to problems with particular VPA implementations rather than with the approach itself. However, it was noticed in simulation trials involving hyperstable y_t ($y_t = qB_t^\beta$, $\beta \ll 1.0$) that SCA estimates are much less prone to upward bias than VPA estimates (i.e., SCA is more robust to violations of the $y = qB$ proportionality assumption) providing the v_a are stable over time, apparently because SCA "sees" information about exploitation rates (and hence stock sizes) in the catch-at-age data.

A more important implication is that we can develop reasonably precise simulations of estimation performance for a given age-structured assessment procedure, without carrying out the tedious calculations of reapplying the whole procedure in each simulation year. The key to whether we can do this for a given case, and use eq. 1 to predict $B_{t|t}$, is whether we can bypass the $B_{t|t-1}$ prediction using eq. 4 (which requires the full estimation results for $N_{a,t|t-1}$) by using a simplified calculation like eqs. 5–6 that requires only $B_{t-1|t-1}$ and the recruitment forecasting method for $N_{1,t|t-1}$. It is easy to check the accuracy of such simplifications using the same simulations of estimation performance as needed to obtain regression estimates of effective Kalman gains K^*; it is particularly important to carry out such checks in cases where the vulnerability schedule v_a may be dome-shaped or may change substantially over time (changes in age at recruitment k).

An example of using eqs. 1–5 in a context where it is impractical to simulate full age-structured assessments is the Ecopath with Ecosim (EwE) software (www.ecopath.org), where we provide options for closed-loop simulation of multispecies, multigear assessment and harvest regulation. The aim in those simulations is to demonstrate how monitoring and assessment errors may impact ability to achieve ecosystem-scale harvest management goals.

Propagation of biomass estimation errors

Results of the previous section can be used to develop an analytical model for patterns of change over time in biomass estimation errors $\xi_t = B_{t|t} - B_t$. Factors that cause persistent positive ξ_t should be a major concern in stock assessment as such factors can lead to severe overfishing in situations

where fishers can achieve TACs set on the basis of $B_{t|t}$ even when B_t has declined severely, as for example in case of Newfoundland northern cod (Walters and McGuire 1996).

Consider situations where enough historical data have been gathered to result in stable Kalman gains K^* (5–10 years of data), where eq. 5 can be used to predict $B_{t|t-1}$ without substantial inaccuracy in the growth–survival factor g_t (i.e., where prediction errors are dominated by uncertainties about $B_{t-1|t-1}$ and recruitment, by far the most common situations), and where observation errors ε_t in the observation model $y_t q B_t + \varepsilon_t$ are random and independent over time. Subtracting B_t from both sides of eq. 1 while assuming $K_t^* = K^*$, substituting eq. 5 for $B_{t|t-1}$, and rearranging the resulting equation leads after some tedious algebra to the following relationship:

$$(9) \qquad \xi_t = (1 - K^*)g_{t-1}\xi_{t-1} + \frac{K^*}{q}\varepsilon_t + \left(\frac{q_t}{q} - 1\right)K^* B_t$$
$$+ (1 - K^*)w_k(N_{kt|t-1} - N_{kt})$$

where q is the assumed survey catchability, q_t is the "true" q_t in the relationship $y_t = q_t B_t + \varepsilon_t$ for year t, and $(N_{kt|t-1} - N_{kt})$ is the prediction error for number of age k recruits in year t based on data available up to year $t - 1$ (N_{kt} = true number of recruits in year t). The first term in eq. 9 represents autoregressive effects of past measurement errors on ξ_t. The second term represents random effects on ξ_t due to measurement errors in year t. The third term represents possible effects of trends or density-dependence in q (e.g., hyperstability). The fourth term reflects the importance of recruitment predictions on the assessment error. Note that a considerably more complex equation results when growth–survival effects g_t are highly uncertain.

One way to visualize the predictions of eq. 9 is to think of the biomass estimation errors as having two dynamic components. There is an autocorrelated random component with magnitude maintained by ongoing random observation errors and by random recruitment deviations from predictions based on recent average or stock–recruitment relations. There is a more worrisome, systematic component that "pumps" positive bias into the error each year in cases where q_t is increasing and (or) recruitment predictions are persistently too high. Walters and McGuire (1996) argue that precisely this bias pumping effect occurred in the northern cod assessments, due to both hyperstability in catch per effort (and possibly even survey) y_t indices and to systematic overestimation of recruitment. In VPA assessments, systematic overestimation of recruitment can occur as the result of using average past recruitment both as a predictor during periods of low juvenile survival and also to direct overestimates of recruitment when v_a for younger fish has been increasing (e.g., resulting from fishers targeting smaller fish as abundance of older fish declines).

Discussion

The existence of tight regression relationships between $B_{t|t} - B_{t|t-1}$ and $B_t^* - B_{t|t-1}$ should not be interpreted as evidence that a given assessment procedure is performing well at estimating B_t. To the contrary, such relationships imply

that whatever biases there may be in estimating B_t^* and $B_{t|t-1}$ are likely to persist over time, whether or not many new observations of relative abundance y_t and catch-at-age $C_{a,t}$ are included in each new application of the procedure and whether or not the procedure involves estimation of many nuisance parameters to account for historical variability.

An interesting point about recruitment variation in the simulations was that such increases in variability, e.g., from lognormal standard deviations of 0.6 to 1.0, certainly caused increased prediction errors $(B_{t|t-1} - B_t)$ but did not cause increased variation around the prediction relationships based on Kalman filter equations. The regression estimates of Kalman filter weights K^* increase as recruitment variance increases, as they should as $B_{t|t-1}$ becomes a less reliable predictor of B_t, but there remains a nearly linear effect of new observations y_t on updating the B_t estimates.

It is disturbing, and even a bit humiliating, to see that the complicated equations, expensive data, and hard work that typically go into an age-structured stock assessment are likely to produce results that can be mimicked by almost trivially simple relationships such as eqs. 1 and 5. But the existence of such relationships does not imply that complex assessments are not worthwhile; they need to be done in order to estimate the "empirical" gain parameters K_t^* and also to estimate other parameters (besides $B_{t|t}$) that are needed for management, including vulnerability schedules and recruitment time series.

Although the K_t^* version of eq. 1 was derived by thinking of "classic" stock assessment data (catches at age along with an abundance trend index) that permit calculation of B_t^* as y_t/q_{t-1}, it can be used to predict behavior of any procedure that combines prediction of stock based on past data $(B_{t|t-1})$ with some new direct estimate B_t^* based on data gathered in year t. An example would be to calculate the new direct estimate as $B_t^* = H_t/u_t^*$, where u_t^* is a direct estimate of the exploitation rate in year t based on a short-term tagging experiment. Martell and Walters (2001) argue that such estimators could potentially lead to much cheaper and more accurate assessment methods than those based on classic data, especially in view of the high costs and risks of trying to obtain survey y_t that are actually proportional to abundance. Another possibility would be to calculate u_t^* using a swept-area method based on careful analysis of spatial logbook data.

There is an unfortunate and very dangerous interaction between the autoregressive and bias pumping components of eq. 9 in management cases where positive estimation errors will result in setting TACs higher than should be allowed. In such cases, positive ξ_t might be caused initially by a few years of unexpectedly low recruitments (positive values for fourth term of eq. 9). The high TACs resulting from these errors will then drive the stock downward, which will lead to increased risk of overly optimistic predictions of recruitment and also of q_t/q increasing due to range contraction effects even if the y_t represent survey rather than catch-per-effort data. This vicious circle of positive ξ_t leading to overharvest that in turn causes increasing bias pumping of ξ_t and still further overharvesting is perhaps the greatest danger that stock assessment scientists should consider in critical analysis of possible biases in stock assessment procedures.

Acknowledgements

References

Butterworth, D.S., Cochrane, K.L., and DeOliviera, J.A.A. 1997. Management procedures: a better way to manage fisheries? The South African experience. Am. Fish. Soc. Symp. **20**: 83–90.

Cooke, J.G. 1999. Improvement of fishery-management advice through simulation testing of harvest algorithms. ICES J. Mar. Sci. **56**: 797–810.

Gelb, A. (*Editor*). 1974. Applied optimal estimation. MIT Press, Cambridge, Mass.

Martell, S., and Walters, C. 2001. Implementing harvest rate strategies by directly monitoring exploitation rates and estimating catchability changes. Bull. Mar. Sci. **70**: 695–713.

NRC (National Research Council). 1998. Improving fish stock assessments. National Academy Press, Washington, D.C.

Pella, J. 1993. Utility of structural time series models and the Kalman filter for predicting the consequences of fishery actions. *In* Proceedings of the International Symposium on Management Strategies for Exploited Fish Populations. *Edited by* G. Kruse. Alaska Sea Grant College Program Report No. 93-02, Fairbanks, Alaska. pp. 571–593.

Punt, A.E., and Smith, A.D.M. 1999. Harvest strategy evaluation for the eastern stock of gemfish (*Rexea solandri*). ICES J. Mar. Sci. **56**: 860–875.

Quinn, T.J., and Deriso, R. 1999. Quantitative fish dynamics. Oxford University Press, New York.

Reed, W.J., and Simons, C.M. 1996. Analyzing catch–effort data by means of the Kalman filter. Can. J. Fish. Aquat. Sci. **53**: 2157–2166.

Walters, C. 1986. Adaptive management of renewable resources. McMillan Publishing Co., New York.

Walters, C.J., and McGuire, J.J. 1996. Lessons for stock assessment from the northern cod collapse. Rev. Fish Biol. Fish. **6**: 125–137.

Walters, C.J., and Punt, A. 1994. Placing odds on sustainable catch using virtual population analysis and survey data. Can. J. Fish. Aquat. Sci. **51**: 946–958.