International Council for the Exploration of the Sea

CM 2007/O:04
Theme Session O: Flying outside the ICES Assessment WG paradigm

A review of Fishery-Independent assessment models, and initial evaluation based on simulated data

John Cotter, Rob Fryer, Benoit Mesnil, Coby Needle, Dankert Skagen, Maria-Teresa Spedicato, Verena Trenkel

Abstract

Large uncertainties in the catch data (official landings and discards) are undermining ICES' ability to provide valid management advice based on the conventional approach of analytical assessments. There is thus an urgent need to consider alternative tools that do not depend on long series of precise catches, with their age composition. This paper presents a few fishery-independent assessment models developed by the EU project FISBOAT (Fishery Independent Survey Based Operational Assessment Tools). It also reports on rudimentary tests based on simulated data, following the same protocol as an evaluation study conducted by the US National Research Council in 1997. It appears that the survey-based assessment models at hand are able to reliably capture the major signal in biomass and recruitment, although they smooth out transient changes. However, they cannot provide absolute abundance estimates, but only relative values on an arbitrary scale. Their operationalisation in ICES would thus require an adaptation of the advisory framework, in terms of nature of the advice and definition of reference points; indeed, this might be needed anyway, if we were more lucid about the myth of VPA estimates being absolute. It is also noted that survey-based approaches have the potential to provide much more rapid updates of the state of stocks than catch-based methods.

Contact author: Benoit Mesnil; Département Ecologie et Modèles pour l'Halieutique; IFREMER, B.P. 21105, Nantes Cedex 03, France. [tel. +33 240 37 40 09; fax +33 240 37 40 75, e-mail Benoit.Mesnil@ifremer.fr]

Introduction

All stock assessment methods (whether they involve surplus-production, delay-difference, stock-reduction, Collie-Sissenwine or analytical dynamic pool models) used by scientific organisations to advise fisheries managers on the state of fish stocks require a knowledge of total catches to estimate the model parameters and other quantities of management interest. Errors in the input catch figures translate directly into similar errors in stock abundance estimates (e.g., Quinn and Deriso, 1999), and if their magnitude varies from year to year the assessments may not even reflect the relative changes in the state of the resources. When catch is also the support of management control, like in TAC systems, there is often a temptation for fishers or states to mis-report for tactical reasons, especially when catch quotas become very restrictive. This has been the case in Europe in recent years and for over a decade ICES has repeatedly stated that the deterioration of the catch data was threatening its ability to provide managers with the type of advice they require to run the current policies.

In the face of this threat, the European Commission has identified the development of operational fishery-independent (i.e. catch-free) assessment tools as one of the priority topics of research in support of the Common Fisheries Policy in the 2003 round of calls for scientific projects. This has been taken up by the FISBOAT (Fishery Independent Survey Based Operational Assessment Tools) project consortium, with participants from 11 research institutes, which completed its work by mid-2007. In this paper we only report on the findings of one work package which was tasked to "*supply methods for analysing fishery independent stock assessment data to provide managers with relevant information about the stock and its exploitation*". Methods are here understood to mean both the mathematical models and the procedures to estimate their parameters.

Section 1 provides a concise overview of the six fishery-independent (F-I) stock assessment methods that were specifically developed, or elaborated upon for use without catch data, during the project. The models are presented with comments on parameter estimation issues, and practical guidelines or caveats regarding their use in assessment and advisory groups are provided.

One final products of the FISBOAT project is a full evaluation of the survey based methods through a simulation-testing evaluation framework (operating model, harvest rule, etc.), but some elements are missing to run that at the moment. However, in order to gain some early understanding of the capabilities of the F-I models, the group had decided to carry out a simpler testing exercise using artificial data with known properties, following the same protocol as an evaluation study conducted in the USA on catch-based (mostly age-structured) assessment models (NRC, 1998). Sections 2 and 3 recount the conditions and results of these preliminary probing tests carried out on four of the six models. Section 4 concludes on the insight gained during the project into the potential performance of the F-I methods for assessment of stock status, and on some implications for the European (ICES) advisory system.

1. Methods considered

The F-I methods developed, or adapted, for the FISBOAT project fall into two categories:
1) methods intended to estimate abundance, or trends thereof; this group includes stage-structured BREM, age-structured SURBA, TSA and YCC, and length-structured LENSUR;
2) simulation methods, to assess the effects of changes in biological or management parameters, represented by ALADYM.
Key features of each method are described in this section, starting with estimation methods. A summary categorisation of the methods with regards to data needs and estimation approach is provided in Table 1. The computer codes and documentation are available on the FISBOAT website.

1.1. Biomass random effects model (BREM)[1]

- Model description

The population dynamics is formulated as the difference model from Hilborn and Walters (1992, p. 336):

$$B_t = R_t + g_t B_{t-1} \qquad (1)$$

where B_t is the total population biomass, R_t the recruitment in biomass in year t-1 and g_t the net biomass growth rate, which is the balance between individual growth and total (natural + fishing) mortality. Recruitment is assumed to follow a logNormal distribution without any stock-recruitment relationship:

$$\log(R_t) \sim N(\mu, \sigma_R^2) \qquad (2)$$

Biomass growth is modelled by a random walk on the log-scale, to reflect the assumption that Z, which is part of g, does not vary wildly from year to year:

$$\log(g_t) = \log(g_{t-1}) + \varepsilon_t^g \text{ with } \varepsilon_t^g \sim N(0, \sigma_g^2). \qquad (3)$$

$$g_t = g_{t-1} \exp(\varepsilon_t^g)$$

Thus both recruitment R_t and biomass growth g_t are treated as random effects with parameters μ and σ_R^2, and g_1 (t=1) and σ_g^2 respectively.

The observation model has two components. The first one is for an index of total biomass at time t (recruits included) and the second for an index of recruits only. Both are assumed to follow logNormal distributions with common variance and catchability coefficient:

$$\log(IB_t) \sim N(\log(q_b B_t), \sigma_{ib}^2) \qquad (4)$$

$$\log(IR_t) \sim N(\log(q_r R_t), \sigma_{ir}^2). \qquad (5)$$

In order to ensure identifiability, the following constraints are imposed: $q_b = 1$ and $\sigma_{ib}^2 = \sigma_{ir}^2$.

- Sensitivity and robustness issues

Convergence of the parameter estimation algorithm depends critically on sensible starting values. The above mentioned constraints allow parameter identifiability, but the effect of setting $q_b = 1$ is that biomass estimates can only be relative not absolute. In addition, the estimates of recruitment and catchability for recruits q_r are confounded to some degree. This appears as strong correlation between estimates.

- Input and Output

BREM only requires two series of survey indices in mass, one for the total population (adults + recruits) and one for the recruits alone; splitting out the recruits can be based on age readings but there are favourable cases where a reasonable cut-off size may be identified by inspection of the length compositions. Note that knowledge of M is not required, and that occasional gaps in survey series are not likely to affect the estimation. An extension handling two series of indices per category (e.g. acoustic and egg surveys) has been developed (Trenkel, 2006, 2007).

---

[1] Contributed by Verena Trenkel, Ifremer, France

Seven parameters are estimated: B_1 (biomass in year 1), g_1 (biomass growth in year 1), $\log(\sigma_g)$ (standard deviation of growth), μ (mean recruitment for base normal), $\log(\sigma_R)$ (standard deviation of recruitment for base normal), q_r (catchability of recruits) and σ_i (standard deviation of observation error for base normal). Plugging converged estimates into Eq. 1 yields estimated time trajectories of relative total biomass and annual recruitment. In addition, standard deviations are available for biomass estimates, but NOT for recruitment estimates as these are random effects, not real parameters.

- Implementation issues
Parameter estimation by maximum likelihood is implemented in AD model builder (Fournier 2005) using the random effects module. Run time for NRC set 1 was about 20 sec. Note that run time does not increase with the number of years. Rather it depends on how good the starting values are.

- Miscellaneous comments
Future recruitment could be predicted using the fitted logNormal distribution, either as expected recruitment or by drawing a random recruitment value from the distribution. The relationship between model predictions and commercial quantities is not obvious.

1.2. LENSUR[2]

- Model description
Lensur is a newly written program for assessing a stock with only length-structured data. *Lensur* has an operating model that generates an artificial population in numbers by length class and time step, as specified by a set of parameters. Model observations are derived from the operating model in an observation model, and parameters are estimated by minimising the deviation of the model observations from real observations, as expressed by an objective function. The objective function so far is a sum of squared log residuals. A minimisation routine searches over the space of parameters and calls the objective function for each parameter set, to find the parameter set that gives the best value, i.e. the best model fit. This is regarded as the estimate of the population according to the data. This places the method within the framework of ‘statistical models’, where in this case the population is constructed so that stock numbers are represented by length.

The method used to obtain length distributions in the population is to follow an ensemble of trajectories representing ‘super-individuals’ over time, each with its own growth characteristics, and time and length at entry (a 'Lagrangian' approach). Hence, internally the population is represented both by length according to a growth model, and age, represented by the time that has passed for each trajectory since it entered the populations. Each trajectory enters the population at a randomly drawn time with a certain number of fish and randomly drawn growth parameters. The abundance and length of each trajectory is calculated for each time step. The whole population is the sum of all trajectories.

- Sensitivity and robustness issues
The implementation in Fisboat further restricts the data to survey data only. This implies strong limitations on what can be inferred from the data in terms of population abundance and exploitation rates, and the method has so far primarily been used to study these limitations (see miscellaneous comments below).

- Input and Output
Operating model parameters are initial numbers at length, numbers (recruits) entering the population each year, growth parameters (k and Linf), selection at length in the fishery, annual

[2] Contributed by Dankert Skagen, IMR, Bergen, Norway

fishing mortality and natural mortality. Parameters in the observation model are survey catchabilities. These are specified as separable, i.e. with a catchability at length and a year factor. The latter normally is assumed to be constant. The program allows the user to specify, for each parameter, whether this parameter shall be estimated by the optimisation routine, or remain fixed at a given value.

Catches at length in each time step are derived from the abundance at length and fishing mortalities at length. These results are not used in the Fisboat framework, but may be useful if catch data (at length or in biomass) are available.

- Implementation issues

Lensur is programmed in Fortran 77, and will be implemented in FLR in the near future. The FLR version is a slightly reduced one, in particular with respect to input-output.
Optimisation is by a searching routine, which is slow, but very robust. This may be an advantage when the method is incorporated in a framework like FLR where there is no user interaction when the program is run.

- Miscellaneous comments

The program can also be used as a data generator, by extracting catches or the output from the survey observations model as artificial data, and numbers from the operating model as the true stock. Noise can be added to the data output, either as random noise, as a random year factor or both. Such data were used in studies on performance. First, it has been confirmed that with no noise in the data, the model fits the data virtually perfectly, and with the right population numbers. Further exploration of the model performance has concentrated on the limitations in what can be inferred with these sparse data. Theoretical considerations and the experience gained using the model on artificial data, leads to the following conclusions about the method:

- Survey indices at length by themselves carry insufficient information for a full stock assessment. Firstly, all such data are relative, and some additional constraint is needed to scale the stock abundance to absolute values. Furthermore, growth and mortality are confounded in the sense that they influence the length distributions in the surveys in a complementary way. Hence, mortality estimates are conditional on assumptions on growth rate. Finally, noise in the data is amplified when translated into mortality estimates. Therefore some constraining assumptions have to be made on the mortalities, and the results are conditional on these assumptions.
- The experience so far is that simple smoothing of the indices is clearly insufficient to avoid undue influence of random noise in the data. Applying a penalty on the year-to-year variation in F takes most of the noise away, and combining that with smoothing of the survey indices removes even more of the noise from the results. When the true fishing mortality is variable, such variations become damped, however.
- Given the theoretical limitations outlined above, it appears that by assessing a stock with only survey indices at length in a statistical method with a length disaggregated model population, it is not possible to provide reliable estimates of variations in exploitation. However, estimates of the stock and the level of, and trends in, the exploitation can be achieved conditional on assumptions about growth rate. The estimates will also be conditional on assumptions about trends in exploitation. If these assumptions are realistic, the estimates obtained will be so as well.

Therefore, in a management context, having length disaggregated survey data as the primary source of information about the stock is only likely to work if there is additional information on trends in the exploitation rate. The assessment will then provide the information about trends in stock abundance on a relative scale, which at least in principle can be translated into harvest rules.

1.3. SURBA[3]

- Model description

The basis of *SURBA* is a simple survey-based separable model of mortality. This model was first applied to European research-vessel survey data by Cook (1997, 2004), but it has a long history in catch-based fisheries stock assessment (Pope and Shepherd 1982, Deriso *et al* 1985, Gudmundsson 1986, Johnson and Quinn 1987, Patterson and Melvin 1996; see Quinn and Deriso 1999 for a summary). The separable model used in *SURBA* assumes that total mortality $Z_{a,y}$ for ages *a* and *y* can expressed as:

$$Z_{a,y} = s_a \times f_y,$$

where s_a and f_y are respectively the age and year effects of mortality. Note that this differs from the usual assumption in that total mortality *Z* is the quantity of interest, rather than fishing mortality *F*. Then, given $Z_{a,y}$, abundance $N_{a,y}$ can be derived as:

$$N_{a,y} = r_{y_0} \exp\left(-\sum_{m=a_0}^{a-1} \sum_{n=y_0}^{y-1} Z_{m,n} \right)$$

where a_0 and $y_0 = y - a - a_0$ are respectively the age and year in which the fish measured as $N_{a,y}$ first recruit to the observed population. Thus the abundance at each age and year of a cohort is given by the recruiting abundance r_{y_0} of the relevant cohort modified by the cumulative effect of mortality during its lifetime. Parameters are estimated by minimising the weighted sum-of-squares of observed and estimated abundance indices. All abundance estimates are relative.

This simple basis has been expanded considerably over recent years, as the model has been road-tested in ICES assessment working groups (and elsewhere) and modified where necessary. The development is summarised in Needle (2002b, 2002d, 2003d, 2004a, 2004b) and Beare *et al* (2005), but in brief:

- Index catchabilities and SSQ weightings can both be defined by the user.
- Biomass indices can be used, as well as multiple age-structured indices.
- The year-effect for the last year is set to the mean of the previous three year effects, as the terminal year-effect cannot be determined directly from the data (although work is progressing on improving this estimate; see below).
- Age-structured indices are all back-shifted to the start of the year, using the current estimate of *Z*. This allows them to be compared directly, and ensures firstly, that abundance indices refer to Jan 1, and secondly, that mortality estimates relate to the calendar year rather than the year between successive cruises of a given survey.
- Biomass indices are shifted forwards to spawning time before inclusion in the parameter estimation process.
- Optionally, a smoothing term can be added to the SSQ to penalise excessive inter-annual variation in estimated year effects. The degree of smoothing is determined by a user-defined variable λ.
- The reference age (that is, the age at which the age-effect *s* is fixed to 1.0) can also be defined by the user.
- Estimated variances (and thereby confidence intervals) of mean *Z* and recruitment are derived from the variance-covariance matrix of the model fit, using the delta method. Variances for abundance and SSB are currently being implemented.
- Retrospective runs are generated automatically, with the last year of data being moved back one year at a time until half of the original time-series remains. This facility can be switched off by the user if required.
- A scan facility has recently been added. With this, the user can automatically run assessments with a range of choices for smoothing, the reference age, and catchability on the first age, and evaluate model sensitivity to these essentially *ad hoc* settings.

[3] Contributed by Coby L. Needle, FRS Marine Laboratory, Aberdeen, UK.

Planned future work includes:

- Improving the terminal year-effect estimate. This may be possible if there are two or more surveys at different times of year, in which case the relative decline in indices during the year may give an idea of mortality during the year.
- Implementation of bootstrap uncertainty estimation, via a multivariate parametric bootstrap.
- Improved scanning procedure.
- Restructured input dialogues.
- Inverse-variance SSQ weighting.
- The results of the scan procedure cannot currently be plotted within *SURBA* itself, although an SPLUS script is provided. The plotting procedure needs to be updated to accommodate this.

- Sensitivity and robustness
The model is most sensitive to assumptions about catchability. In particular, estimates of Z can be very different under different assumptions about catchability; SSB estimates are more robust. Z estimates can be very uncertain in any case, and it is not uncommon for there to be no significant evidence of any changes in the levels of Z. However, this may well be true for most models in which uncertainty is estimated. The ICES North Sea Demersal WG encountered difficulties in fitting *SURBA* to flatfish survey data during its 2005 meeting, and these are still not resolved. Finally, the automated scanning routine sometimes fails – values scanned over need to be interactively defined in future.

- Inputs and output
SURBA uses the Lowestoft VPA input format, and currently expects to see the full set of such files – which means that dummy catch-based data files had to be set up in order to analyse the NRC datasets. The inputs that are actually required for fitting the model are age-structured tuning indices, and (optionally) biomass tuning indices. The user can also define catchability and SSQ weightings for both types of index, along with values for the smoother λ and the reference age.

Both text and graphical outputs are provided by the program. Text outputs include parameter estimates with variances, mortality and relative abundance estimates, estimated variances for mean Z and recruitment, log residuals, stock summaries (SSB etc.), results of retrospective and scan runs, and goodness-of-fit statistics. Plots include exploratory raw-data figures (such as catch curves), model fits and stock summaries, residuals, and retrospective summaries.

- Implementation issues
SURBA (currently Version 3.0) is implemented in a Windows user interface, in which diagnostic plots are automatically generated. The run time for NRC set 1 was 6 s (standard), 40 s (standard + 15 retrospective runs), and 7 m 47 s (105-run scan).

- Predictive ability
SURBA does not currently feature a forecasting mode, although this is planned in the near future. It is intended that this will roll forwards the population from different starting points arising from the bootstrap runs mentioned above, leading to stochastic forecasts. This will need assumptions about weights, exploitation, and recruitment.

- Relation to management indicators
Abundance estimates (and therefore biomass measures) are currently generated by *SURBA* on a relative scale only, and are usually plotted as mean-standardised values for ease of comparison. Furthermore, *SURBA* provides estimates of total mortality Z rather than fishing mortality F (although, given the tentative nature of most natural mortality estimates, this is true of catch-at-

age methods also). Therefore *SURBA* can be used to provide advice on relative trends in abundance and total mortality, but not absolute levels. It is possible to generate pseudo-absolute abundance estimates by using a catch-at-age VPA to estimate survey catchabilities-at-age using data from some period in the past, and then applying these to recent *SURBA*-derived relative population estimates to scale them up to a level commensurate with that indicated by catch data (Needle 2004a). However, this requires assumptions that there was a period when catch data were reliable, and that the relationship between survey and fishery catchability has remained constant ever since, and these can be hard to maintain. It is also possible, of course, to produce F estimates by subtracting fixed M values from the Z estimates produced by *SURBA*.

If *SURBA* (or any other survey-based approach) is to be used as a management tool, there needs to be a clear idea of the management framework in which such a tool would be used. In other words, reference points for mortality and biomass would need to be redefined on the basis of total mortality and relative biomass, respectively.

1.4. Time series analysis (TSA)[4]

- Model description

TSA, or 'Time Series Analysis', is a state space framework for modelling a fishery. The initial implementation, by Gudmundsson (1994), modelled commercial catch-at-age data, with survey indices-at-age used as auxiliary information. Here, the framework is adapted to model the indices-at-age from a single survey. The state equations relate the log numbers-at-age and fishing mortalities-at-age in year y to those in year y–1. Log numbers-at-age in year y are given by:

$$n(a,y) = n(a-1,y-1) - Z(a-1,y-1) \quad a > 1$$

$$n(1,y) = \mu + \mathrm{NID}(0,\sigma^2_{\mathrm{recruit}})$$

where NID stands for Normal Independent Deviate. Fishing mortalities evolve according to the following model:

$$\log F(a,y) = U(a,y) + V(y)$$

$$U(a,y) = U(a,y-1) + \mathrm{NID}(0,\sigma^2_U) \quad \text{with the constraint that } \textstyle\sum_a U(a,y) = 0$$

$$V(y) = Y(y) + \mathrm{NID}(0,\sigma^2_V)$$

$$Y(y) = Y(y-1) + \mathrm{NID}(0,\sigma^2_Y)$$

Thus, log fishing mortality is separated into an age component $U(a,y)$ and a year component $V(y)$, both of which can evolve over time. Finally, the state vector consists of the $n(a,y)$, log $F(a,y)$, $U(a,y)$, $V(y)$ and $Y(y)$.

The observation equations are given by:

$$i(a,y) = q(a) + n(a,y) + \varepsilon(a,y)$$

where $i(a,y)$ are the log indices-at-age, $q(a)$ are the survey log catchabilities, and the $\varepsilon(a,y)$ are assumed to be NID with zero mean and standard deviation $\sigma_{\mathrm{survey}}\ \lambda(a)\ \delta(a,y)$. The $\lambda(a)$ are initially taken to be unity, but can be adjusted later if the errors associated with some ages are larger than for others. The $\delta(a,y)$ are also initially taken to be unity, but can be inflated to decrease the influence of outliers. It is assumed that the survey takes place at the start of the year.

The model is fitted using the Kalman Filter, with the parameters μ, $\sigma_{\mathrm{recruit}}$, σ_{survey}, σ_U, σ_V, σ_Y, $q(a)$, $U(a,1)$ estimated by maximum likelihood. For identifiability, $q(1)$, $V(1)$ are taken to be zero. For stability, some constraints must be put on the $q(a)$: the current implementation takes the $q(a)$, $a > 1$, to change linearly with age.

[4] Contributed by Rob Fryer, FRS Marine Laboratory, Aberdeen, UK.

- Sensitivity / robustness

Good starting values can be difficult to find for a new stock: some iteration and experience is required. Poor starting values will either make the current implementation crash or will erroneously suggest that the starting values are optimal. However, once good starting values have been found, the implementation is robust to the addition of an extra year's data, etc. The same starting values were used for all the NRC data sets (except for the clean set).

The method works on the log scale, so zero indices must be replaced by some small positive value. Unity was used for the NRC data sets. This means that the method can only be sensibly applied to those age classes where zero indices do not often occur – typically the younger age classes. An option would be to group older age classes into a single plus group, but this has not been implemented yet.

Very large year classes can cause a problem, because they can unduly dominate the parameter estimates associated with recruitment (i.e. μ and $\sigma_{recruit}$). It is possible to reduce their impact on these estimates, but this is done manually following graphical inspection of standardised prediction errors.

- Inputs and outputs

Inputs:

- survey indices-at-age; can handle several surveys in sequence (e.g. if there is a change in q in the single survey at a particular time), but not in parallel. Missing survey data are accepted, at the cost of increased standard errors in estimates around missing years while not causing bias.
- if natural mortalities-at-age are provided, then (relative) fishing mortalities will be estimated, otherwise (relative) total mortality Z will be estimated.

Outputs:

- estimates of relative numbers-at-age with approximate coefficients of variation; these can not be combined across age classes (there is a separate scaling factor for each age class), so it is not possible to estimate (relative) biomass, etc; however, sensible proxies for stock biomass can be estimated;
- estimates of relative fishing mortalities-at-age with approximate coefficients of variation; these can be combined across age classes, so it is possible to estimate (relative) mean fishing mortalities for groups of age classes;
- evidence of persistent changes in fishing mortality, either overall, or as departures from separability.

- Implementation

- Fortran 90, using NAG routines.
- Took ~ 30 seconds to run NRC set 1 on a 1.8GHz, 524MB RAM laptop.

- Predictive ability

The method can predict both relative numbers-at-age and fishing mortalities-at-age (with approximate coefficients of variation) as far into the future as required.

- Relationship with commercial quantities

If natural mortalities-at-age are provided, then relative fishing mortalities-at-age are estimated (otherwise only relative total mortalities-at-age Z are estimated).

1.5. Year-class curve (YCC) method[5]

- Model description

A 'year-class' curve is a plot of log CPUE over age for a single year-class of a species. Marine fish caught in trawls typically show nearly linear year-class curves for ages that are fully selected. The usual model of mortality over time *t*, assuming no net migration to or from the stock, is

$$\frac{dN}{dt} = ZN$$

where *Z*, the instantaneous rate of total mortality, is here expected to have a negative value. [The absence of a minus sign before *Z* is unconventional in fisheries work but leads to Equation (2) having all terms positive, as is conventional for regression models.] Solving gives

$$N_t = N_0 \exp(Zt). \qquad (1)$$

We now assume that catch per unit effort (cpue, denoted *U*) is a constant proportion of *N*, i.e. $U = qN$ for all ages included in the analysis, and that *Z* represents a constant, average value over time. Then, taking natural logarithms of Equation (1), restricting attention to one year-class, *c*, substituting *age* for *t*, and adding a random error term, *e*, gives the basic model for a year-class curve:

$$\log U_{a,c} = \log(U_{0,c}) + Z\,age + e_{a,c} \qquad (2)$$

where $U_{0,c}$ is the cpue (or survey) index for age zero, *a* is the age-class, i.e. the age in years as an integer index, while *age* is age in years as a real number. *e* is assumed to be normally distributed around zero with residual variance σ_e^2. Additional linear terms may be added to equation (2) to allow for varying selectivity of the survey trawl with age, to allow for different RV (or commercial) fleets having different catchabilities, *q*, and to allow for gradual changes in *Z* over time. The latter is achieved using polynomials in *age* and *year* with a minimum of additional parameters so as to yield best precision of estimation with the available observed data.

Different series of cpue data are likely to estimate year-class curves with different precision depending on the season and area covered by the fleet, on the precision of age-reading and other practical aspects, and on how well the chosen model fits the data. Weighting of different data sets to reflect their precision with respect to the chosen model is therefore desirable. Cotter and Buckland (2004) suggest that the weighting estimated for each fleet's data set should be balanced with the reciprocal of the estimated residual variance specific to that fleet computed after the model is fitted, i.e. $\hat{w}_f \propto \hat{\sigma}_f^{-2}$. They describe how the method can be implemented using iteratively weighted least squares (IWLS) taking into account the d.o.f. contributed by each fleet to the estimates of each parameter. Usually, 2 or 3 iterations produce stable values. Additionally, using the fleet specific residual variances, the relative precision of the different fleets can be compared using *F* tests (Cotter 2001). Note that biased cpue series will produce biased weights (Quinn and Deriso 1999, p. 353). Fleets that appear exceptionally precise should be scrutinised to see whether biased sampling may be the cause, e.g. due to clustering of observations in restricted times or places (Cotter and Buckland 2004).

A year-class curve can be fitted repeatedly in a process called *forward validation* that is designed to find the most reliable model for predicting next year's cpue. Starting from an early year and proceeding forwards in the time-series, it finds the differences between the predicted log cpue and the observed log cpue for one year after the time domain of the data used to fit the model. The preferred model is the one whose mean difference is closest to zero, and for which the mean square of the differences is lowest. This is merely a simulation of a fish stock assessment working group making predictions each year for the coming year, then checking

[5] Contributed by John Cotter, CEFAS, UK.

them when the outcome is known. Full details of available models, fleet weighting, and forward validation to find the preferred model are given by Cotter et al. (2007).

- Sensitivity / robustness

Catchability must be constant over time but may vary among different surveys or fleets since intercalibration factors are automatically fitted if required. Changes to the design of an RV survey that might cause a change in catchability (e.g. a different vessel or gear) can be accommodated simply by treating it as a new fleet and fitting an extra intercalibration factor.

Only gradual changes of *Z* are allowed by using polynomials in *year* to a maximum degree of 3. This is intended to minimise the dangers of erroneously treating random measurement errors as trends in the year-class signal over time. However, if sudden, real changes in *Z* actually do occur from year to year, they might be overlooked.

Year-class curves can be fitted across fleets, or nested within. Over- and under-fitting can both caused biased estimates of parameters. Forward validation helps to eliminate such models because they tend to be poor at predicting beyond the observed domain. The AIC may also be used to help with finding the best model.

- Inputs and outputs

The basic input is a standard VPA-type tuning file (Darby and Flatman 1994). *YCC* software operates on a flat file having fleet, age, year, time-of-year, cpue, etc., so such a file may be utilised directly if preferred. Year-class curves are available as plots over time, one per year class. These allow the fitted model to be compared to the observed values to check that the fit is credible. Relative recruitments, and *Z* over age by fleet are also given, along with various other outputs.

- Implementation

Software to fit year-class curves with all the options described here is called *YCC*; it is written in R. The user is asked what terms are wanted in the model, whether terms are to be nested in fleets, whether to switch weighting on or off, whether to use forward validation, and about outputs. The latter may be obtained on screen, as text files, or as graphics. Diagnostics include prediction and residual errors over time, age, and year class. Some of these outputs are illustrated by Cotter et al. (2007). Run times are usually seconds but may increase to a minute or more when there are many fleets, iterative re-weighting, and a long period of forward validation. The model may fail to fit if there are more parameters than observed vectors of cpue-at-age. Missing values may either be omitted from the data set or coded as negative cpue.

- Predictive abilities

Predictions one year ahead of observed data is carried out routinely with forward validation. *YCC* produces tables of predicted cpue-at-age for the year after the final observed year together with prediction mean square errors.

- Relationship with commercial quantities/management indicators

Predicted cpue-at-age in terms of numbers may be converted to weights per unit of effort-at-age using a matrix of weights-at-age by year. These may in turn be converted to spawning stock biomass per unit of effort-at-age using a matrix of maturity-at-age by year. The software allows users to insert independent observed values for each year, if available. Year-class curves fitted to cpue for commercial fleets could provide predicted catches from predicted cpue multiplied by predicted commercial effort under different fishing scenarios. However, *YCC* offers no prediction of next year's recruiting year class.

Z is estimated numerically for each age and year from fitted curves (rather than from fitted parameters which may be individually biased, depending on the model). No assumptions are made about natural mortality, *M*. Fishing mortality, *F*, could be estimated if they were.

1.6 ALADYM simulation method[6]

- Model description

ALADYM (**A**ge-**L**ength B**a**sed **Dy**namic **M**odel) is an age-length based simulation model developed in the conceptual framework of dynamic pool models, following the predictive Thompson & Bell (1934) approach. The model is designed to predict, through simulations, the effects of different fishing pressure scenarios on a single population, in terms of different metrics and indicators. Removals are simulated on the basis of the total mortality rate modulated using harvesting pattern and a fishing activity coefficient. *Aladym* can work in absence of fishery-dependent data, although its predictive capability of real catch levels can be verified using information on commercial catches or fishing activity per month.

From the *Aladym* core model three complementary, but independent, tools have been derived:

A) the quasi-deterministic dynamic tool named *Aladym-r*;

B) the tuning tool *Aladym-z*;

C) the stochastic dynamic tool named *Aladym-q*.

Aladym-q adds to the same mathematical model of *Aladym-r* the capability to deal with the stochastic representation, modelling the uncertainty of estimates related to recruitment, growth and maturity through stochastic processes. This makes *Aladym-q* more suitable for estimating the probability associated to predicted metrics, indicators and reference points. *Aladym-z* has been developed as a specific tool, which starting from the observed values of Z and the description of the life and population traits is able to calculate values of total mortality which better approximates a given scenario.

The model is designed to simulate population dynamics of a given population accounting for differences by sex in growth, maturity and mortality. All the quantities are calculated as vectors with an associated time step Δt (time slice=1 month).

The population dynamics is formulated following the simultaneous evolution of several cohorts at month scale through the exponential population decline model, both in absence (1) and in presence (2) of fishing mortality:

$$\frac{dN}{dt} = -MN \qquad (1)$$

$$\frac{dN}{dt} = -ZN \qquad (2)$$

used respectively in the form (3) and (4):

$$N_{(t+\Delta t),j} = N_{t,j} e^{-M_{t,j} * \Delta t} \qquad (3)$$

$$N_{(t+\Delta t),j} = N_{t,j} e^{-(F_{t,j}+M_{t,j}) * \Delta t} \qquad (4)$$

where j indicates the cohort, t the time, Z, M and F the total, natural and fishing mortality respectively.

Initial numbers in the population are from estimates of recruitment independently obtained (e.g. from trawl surveys). The number of recruits entering the population in successive years can be a vector or is estimated from a stock-recruitment relationship (Beverton & Holt, 1957; Ricker, 1975; Shepherd, 1982; Barrowman & Myers, 2000), with random variations. The number of the events (on monthly basis) generating the offsprings is an input of the model.

The growth process is assumed according to a VBGF and a length-weight relationship; the maturity follows an ogive model.

The natural mortality can be constant for each age/length or a vector by age/length calculated outside the model. Alternatively, it is estimated inside the model from the Chen and Watanabe equations (1989).

The fishing mortality rate $F(L)$ is modelled for each cohort using the following general equation (Sparre and Venema, 1998):

$$F(L) = F_{max} \cdot S(\overline{L})$$

[6] Contributed by Maria Teresa Spedicato, COISPA, Bari, Italy

where F_{max} is the maximum fishing mortality and $S(\overline{L})$ the proportion of retained fish.
In *Aladym* the fishing mortality rate is calculated as follows:

$$F(L) = F_{max} \cdot S(\overline{L}) \cdot f_{act}$$

where maximum fishing mortality (F_{max}) is calculated as follows:

$$F_{max} = QZ_{input} - M_{\min}$$

using the input values of QZ (a Z proxy) and where M_{min} represents the minimum value that the M vector assumes. In addition, a fishing activity coefficient (f_{act}) is introduced in order to consider the possibility of a fishing ban or changes in fishing effort throughout time.
The value of QZ by sex can be assumed, as a first order approximation, numerically equal to the value of Z observed that is obtained from estimations outside the simulation model (e.g. from trawl-survey). A better approximation of QZ is obtained using the tool *Aladym-z*.

- Inputs

- von Bertalanffy growth parameters by sex with associated variability,
- length-weight relationship parameters by sex;
- maturity ogive parameters by sex ($L_{m50\%}$ and $L_{m25\%}$-$L_{m75\%}$ range);
- natural mortality by sex (a constant value or a vector);
- seed values (minimum, maximum, *ln*-mean and *ln*-standard deviation) of recruitment by sex;
- proportion of offsprings entering in the stock by month;
- stock-recruitment relationship parameters or a vector of recruit numbers by month both with associated variability;
- time elapsing from spawning to birth;
- sex-ratio (female/total) of offsprings;
- F_{max} by month or from the model;
- *QZ* by sex;
- selection ogive parameters (2 options) of the gear used by the fleet ($L_{50\%}$ and $L_{25\%}$-$L_{75\%}$ range, $D_{50\%}$ in case of the selectivity option 2);
- fishing activity coefficient by month (0, in case of absence of fishing activity).

In *Aladym-q* the following inputs are also provided:

- the number of realizations;
- the parameters of the defined *pdfs*.

- Harvest control rules

The simulation approach can be used as a tool to convert survey biological information and relative assessment into quantitative HCRs. The options implemented in the simulation model are based on the following aspects: total mortality, gear selectivity (size at first capture $L_{50\%}$ and selection range) and fishing activity (alone or in combination). These three are inputs that can be used to simulate different exploitation scenarios. The effects of HCRs (selectivity and fishing activity) are then analysed in terms of sustainability for the population in the long-term. For example, the ratio between the mean spawning stock biomass and the mean unexploited spawning stock biomass (*SSB/USSB*, output) is also estimated for each harvesting scenario.
A vector of yield (*Y*) by time is also simulated, estimating the catch (*C*) according to the following general equation (Gulland, 1969):

$$C_{\Delta t} = \int_0^{\Delta t} F \cdot N_0 \cdot e^{-Z \cdot \tau} d\tau = \frac{F}{Z} N_0 \cdot \left(1 - e^{-Z \cdot \Delta t}\right)$$

where Δt is the time to which the catch is referred.
Thus the catch (Yield) in the time interval (t, $t+\Delta t$) is computed in *Aladym* as (Sparre and Venema, 1998):

$$Y_{t,j} = \frac{F_{t,j}}{Z_{t,j}} \cdot N_{t,j} \cdot (1 - e^{-(F_{t,j} + M_{t,j}) \cdot \Delta t}) \cdot \mathrm{w}_{\mathrm{age}}$$

- Assumptions and sensitivity

The basic assumptions of the model are:

a) natural mortality as estimated reflects the rate of decline of a population for all causes excluding fishing;
b) total mortality *Z* reliably reflects the decline of ages/sizes in the population, including the effects of different fishing gears;
c) the growth, the natural mortality, and the maturity parameters are assumed constant along the time;
d) given the small time interval (1 month) between cohorts the effect of the spreading of the lengths respect to the ages can be neglected.

The model behaviour is influenced by the consistency between the set of life-history parameters and population dynamics. The model results are thus expected to be particularly sensitive to the stock-recruitment relationship and natural mortality.

To summarise, the main features of the F-I methods reviewed are presented in Tables 1.a,b below.

Table 1. Categorisation of methods

a. Technical

Method	Approach	Estimation	Structure	Input*	Need M?	Output
BREM	Biomass, Random effects	Max-Lik	Two-stage (Rec /Tot), in mass	Indices by stage; 1 or 2 fleets	N	Relative R & Btot
LENSUR	Length, Lagrangian	NLLS	Length	Indices by L. class; 1 fleet; v.B. params		Annual F, relative N@L
SURBA	Separable Z	Weighted NLLS	Age	Age-disagg. indices; n fleets	N	Z, relative R, TSB, SSB & N@age; conf. limits on R & Z
TSA	State-space, Time Series	Kalman filter, Max-Lik	Age	Age-disagg. indices; 1 fleet	N, but can be used (by age & year if avail.)	Relative N@age & Z; Not B
YCC	Y-Class curve	GLS	Age	Age-disagg. indices; n fleets	N	Z, relative R, rel. Bt & SSB, rel. fleets' q, rel. precision of fleets, predicted indices
ALADYM	Simulation	-	Age-Length	v.B. params, Z, selec		multiple

* only those essential for fitting; not for derived quantities such as Btot or SSB

b. Management measures that can be informed

Method	TAC	Effort	Gear/Mesh	Time closure	Other
ALADYM	(y)	Y	Y	Y	
BREM	Y	Y	N	N	
LENSUR	Y	Y	Y		
SURBA	Y	Y	N (unless *M* is known)	N (unless multiple surveys from ≠ times of year)	
TSA	Y	Y			
YCC	Y	Y			

2. Testing procedure

2.1. Data Sets

In absence of a better alternative at the time, we resorted to the suite of data sets concocted for the US National Research Council rounds of tests during 1997. One advantage is that the outcome has been published (NRC, 1998), enabling the performance of other methods to be compared with that of the methods considered by that committee (which all made use of catch and/or catch-at-age data). The data were generated by an age-structured model, where a 15-age population was projected over some 40 years but data for only the last 30 years were retained. Details of the data generation are given in Chapter 5 and Appendix E of the NRC report, and the main features are summarised in Table 2 below. Each data set is a single replication of a combination of stochastic processes[7]. A special comment applies to data set 3, which involves a change in survey vessel (and a near doubling of survey q), a feature that was not explicitly disclosed to the FISBOAT analysts initially and was a clear violation of a basic assumption in their method; however, given the knowledge of a step change in q, all methods are able to deal with this situation and most authors repeated the analysis with each period treated as a distinct survey (run labelled "set 3.2" hereafter), which resulted in improved performance. Also note that data set 5 simulates a case with very low exploitation rate (Yield/Biomass ratio in Table 2).

Table 2. Specifications of the simulated data sets (expanded from NRC 1998).

Set	Population trend	Age at 50% selectivity	Misreporting	Survey q	CV survey q	M	Mean Y/B
1	Depletion	Lower later	0.97-1.03	Constant	0.3	0.18-0.27	0.19
2	Depletion	Lower later	0.68-0.72	Constant	0.3	0.18-0.27	0.12
3	Depletion	Lower later	0.97-1.03	Higher later	0.3	0.18-0.27	0.12
4	Depletion	Constant	0.97-1.03	Constant	0.3	0.18-0.27	0.21
5	Recovery	Constant	0.97-1.03	Constant	0.3	0.18-0.27	0.07
6	2-way trip	Constant	0	Constant	0 (clean set)	0.2	0.15

Since some NRC sets are rather tough, a "clean" set (labelled # 6) was added where survey q has been strictly constant, and indices at age measured without error. This was also generated with an age-structured model comprising 15 age groups, and twenty years of data were output. Methods that break down on this easy set would clearly require some hard work.

The data sets were circulated to methods' authors in advance of a project workshop. The main information that was provided is the matrix of survey indices by age and year. Weights at age, natural mortality (average for the NRC sets, where M varied randomly) and maturity ogive were also provided, in case some methods would need these data, but no information about catches and effort by the fishery was given. It was proposed that analysts focus on the following outputs for comparisons: time series of recruitment (preferably in number); time series of total biomass and, if possible, of total numbers; optionally, time series of SSB.

Clearly these data were not adequate to test length-structured models, such as LENSUR, for which specific test data have to be set up (preferably providing true states, i.e. not on real stocks). The testing framework was also inadequate to evaluate the ALADYM simulation model.

[7] The report of the 2007 Methods WG (ICES CM 2007/RMC:04, Section 2.1.2) may leave the impression that the test data were not corrupted with noise. We point out that the NRC sets 1-5 did include various elements of noise, with perhaps the most relevant for this test being a random logNormal error on the survey indices at age with a 30% CV. Only set 6 was 'clean'.

2.2. Performance metrics

The intention behind selecting the NRC test sets was that comparisons might be possible with the performance achieved by catch-based assessment methods as documented in the NRC report. Since the latter methods are deemed to provide absolute estimates of key management variables, the NRC Committee chose to evaluate the methods based on relative error statistics (i.e. [(estimated – true)/true], both estimates and truth being in absolute value). For F-I methods, however, a clear message from all authors is that these could only provide estimates of relative trends in population variables, and thus the statistics above could not be used. Alternatively, the following approach to a performance metric involving relative values was considered: for each quantity of interest, the time series of estimates, on the one hand, and of true values, on the other hand, are first normalised by subtracting the respective mean and dividing by their SE (years with NA estimates, which are specific to each method, are excluded from both series before computing mean and SE), which gives a common scaling; the mean over years (rather than the sum, to account for NA-related differences in time series' length among methods) of the squared deviations between normalised estimates and normalised truth is computed; the square root of that mean is taken as the summary statistic (kind of RMSE). Although this statistic is not readily interpretable to gauge the performance against standard criteria, it enables fair comparisons between the F-I methods (unfortunately the results of catch-based methods are only shown graphically and not tabulated in the NRC report, otherwise the same statistic could have been computed and both classes of methods compared on equal footing).

The biomass depletion rate, that is the estimate of biomass in the final year divided by that in the first year, as considered in the NRC tests should in principle be the same when based on absolute or relative estimates and was also retained as an indicator for comparisons (for those F-I methods yielding biomass estimates), together with the NRC mild criterion that the relative error compared to the true rate should be within $\pm 25\%$.

As a further aid to compare methods, the estimation CVs for recruitment and biomass (when the method is able to provide them) obtained for each data set were also tabulated.

3. Results of methods comparisons across sets

The relative performance of the F-I methods tested is summarised in Tables 3.a-e for each of the performance metrics described above. Graphical comparisons of the trajectories of estimates vs. the truth (both normalised) are also shown to gain more detailed insight into the behaviour of each method (Figures 1-4).

The first thing to note is that most methods did very well with the clean set #6 (only YCC showed some inconsequential deviations for recruitment estimates), which is reassuring: this validation test indicates that there is no inherent defect in the rationale of these methods, nor in the computer code.

These methods essentially behave as smoothers for noisy indices, and may miss quick transient changes in stock abundance. However, in their expected usage to evaluate "current" stock state by comparing present and historic estimates, none would have caused managers to be misled about the situation of the stock and actions to take in the last decade of the time series. For recruitment, the position of weak or strong year-classes is generally correct, although there are cases of either over-smoothing or over-reaction to the signal in the survey.

Like most VPA tuning methods, these F-I methods make the strong assumption that survey q (by age or stage) is constant over time, and it should not come as a surprise that estimates were badly biased in the tests with set 3.1 where the large step change in q was ignored. In normal circumstances, the assessors would be aware of such marked changes in the survey procedure and would adjust the treatment of their data accordingly, as exemplified by the runs redone as 3.2. Nevertheless, this test highlights the fact that F-I methods are strongly dependent on the quality of the survey, notably the consistency of the survey protocol, as they use no other source of information which might counterbalance poor survey data. In actual life, year-on-year variations in survey design (e.g. due to weather or logistic constraints) or gear rigging are common, and users of F-I methods should be alert that they must take them into account, however benign they may appear at first sight.

In contrast, the test indicates no particular problem with set 5, a case with very low exploitation rate (F << M) which may cause poor convergence of VPA based methods.

Overall, based on inspection of summary statistics and patterns in the plots, all the methods tried in this test perform quite similarly and could be used interchangeably, depending on availability and familiarity with the software. There is a small practical advantage in favour of *BREM* which does not require extensive age compositions. Moreover, *TSA* does not (yet) provide biomass trajectories, and the plots of *SURBA* estimates show occasional wiggliness in some batches of years.

As said earlier, it is not straightforward to compare the performance of the F-I methods with those of the tuned catch-based methods applied to the same data in the NRC tests, since estimates from the latter are not available in tabular form. Coarse comparisons with the biomass trajectories plotted in Appendix I of NRC (1998) indicate that catch-based methods tended to consistently over- or (most often) under-estimate relative to the truth, whereas F-I estimates wander about the true trajectory. Note in passing that with set 5, all catch-based methods under-estimated the true absolute biomass by a considerable amount, but may have preserved the relative trend. More direct, albeit not necessarily easier, comparisons can be made with the estimates of depletion rate for those NRC runs where only the survey data (not the commercial CPUE series not considered here) were used for tuning. F-I methods, notably *BREM*, perform comparatively well and were generally outperformed only by the most highly parameterised catch-based methods.

It must be kept in mind that this evaluation is contingent on, among other things, scenarios where the error in observation of the indices has a CV of 30%, a value which is considered reasonable for well-behaved surveys. If in reality these methods are applied to survey data with larger errors, across the series or in specific years, their reliability in advisory contexts will obviously be poorer.

There is also the limitation that this test is based on a single replication of a stochastic data generation, and that a proper evaluation would require summarising over many replicates – this is the task of another work package in the project. We note, however, that our protocol is the same as the one adopted by an eminent scientific committee.

Table 3. Performance statistics

a. RMS of normalised deviations for Recruits

Method \ Set	1	2	3.1*	3.2*	4	5	Clean
BREM	0.559	0.435	0.775	0.744	0.540	0.548	0.001
SURBA	0.481	0.466	0.752	0.725	0.462	0.495	0.121
TSA	0.556	0.441	0.747		0.486	0.536	0.039
YCC	0.504	0.781	0.621	0.542	0.722	0.461	0.361

* 3.1: set 3 assuming a single consistent survey; 3.2: survey split in two (before/after change in vessel).

b. RMS of normalised deviations for Biomass

Method \ Set	1	2	3.1	3.2	4	5	Clean
BREM	0.207	0.211	0.805	0.524	0.194	0.197	0.012
SURBA	0.402	0.500	0.930	0.892	0.434	0.564	0.031
TSA							
YCC	0.182	0.187	0.869	0.347	0.135	0.146	0.152

c. CV (in %) on Recruits estimates (average over years)

Method \ Set	1	2	3.1	3.2	4	5	Clean
BREM				62.3			
SURBA	18.7	21.7	22.7	15.5	20.7	18.1	3.0
TSA	13.4	18.8	15.9		16.8	16.2	0.05
YCC	44.2	10.5	10.3	15.3	11.0	8.4	23.5

d. CV (in %) on Biomass estimates (average over years)

Method \ Set	1	2	3.1	3.2	4	5	Clean
BREM	46.3	54.6	39.9	69.2	12.1	37.0	14.4
SURBA							
TSA*	9.5	11.7	11.2		11.9	10.9	0.04
YCC							

* CV of GM stock number over ages

e. Relative error (in %) in Depletion rate (Biomass in final year / in year 1)

Results in boldface meet NRC ±25% criterion

Method \ Set	1	2	3.1	3.2	4	5	Clean
BREM	**-22.6**	**-3.9**	193.1	121.2	**15.7**	40.6	**1.0**
SURBA	-30.8	31.4	80.2	77.6	-39.8	**-20.0**	**-5.0**
TSA							
YCC	**-20.1**	42.5	137.9	**-3.5**	**2.0**	32.7	**0.3**

4. Conclusions

Although rudimentary, and awaiting further evaluation in full-fledged management strategy evaluation simulations, this exercise indicates that the F-I methods developed for this project are promising in terms of usefulness and reliability as bases for management advice.

Their main advantage, and indeed their raison-d'être, is that they are not subject to uncertainties in the commercial catches which have caused growing concern and controversies about scientific advice based on VPA approaches in recent years. Moreover, the dependence on catch data is the main reason for the current one-year delay between "data year" and "assessment year", which attracts criticism by managers that response from scientists to their requests is too slow. Clearly, survey-based methods can resolve this timeliness issue, as availability of updated information on stock state is a matter of days after a survey is completed (some overhead is still needed for data auditing, construction of the total area index when this involves more elaborate treatments than just aggregating samples, and mostly for age reading for those F-I methods requiring detailed age compositions). Another bonus with all the methods reviewed here is that their fitting procedures do not require prior knowledge of the natural mortality coefficient, which is a crucial ingredient in many other assessment methods and perhaps the most challenging parameter to estimate (M may still be needed for derived quantities, such as extracting F if management specifically needs it). Finally, it can be seen as an advantage that the methods reviewed have few if any "tuning knobs" to fiddle with.

Evidently, there are a few drawbacks. One is that there is no hope to estimate absolute stock size (overall or for specific ages): all abundance estimates are to be treated as relative, with an arbitrary scaling coefficient (= survey q) between actual and estimated abundance. In itself, this is not necessarily an issue, and examples might easily be found in many areas where decisions of utmost importance to society are made in reaction to relative indicators. The problem with fisheries management in Europe merely arises because, decades ago, scientists successfully sold the idea that they had the skills to deliver advice in absolute terms and the "system" has been built-up on these premises. One consequence is that managers were never educated to make use of alternative flows of information, such as relative indicators coupled with reference points based on past states (if only as a cross-check of the traditional advice), and more seriously that scientists have never formalised and evaluated an advisory process based on such information, although many critics argue that allegedly absolute VPA estimates are effectively relative since they are scaled by input M's which are guessed rather than known. However, this is mostly a problem with the advisory system and it should not count against the performance of the F-I methods *per se*.

A more inherent limitation of F-I methods is that they only use one source of information, and are thus critically dependent on the quality of survey protocols and data. Perceived year-on-year changes in abundance, and ensuing effects on advised management decisions, are likely to be very fragile to inconsistencies in the conduct of surveys (dates, geographical coverage, gear, etc.), and the best professional standards must be adhered to in order to reduce biases. When survey programmes are directed at groups of species (e.g. IBTS), the design tries to achieve a compromise between the needs of various species, and there are often populations whose distribution is only partially covered; this potential bias has to be borne in mind when candidate species are selected for application of F-I methods (and in any case when interpreting the results for advice). Finally, despite the complaint by paymasters that surveys are by far the most costly item in the assessment process, the implication of basing management on F-I approaches may well be that more, rather than less, investment in surveys is required notably for those where the precision of indices is near the limit of acceptability. Although gaps in survey data do not technically impede estimation with the methods reviewed, it is obvious that the quality of assessments degrades quickly when gaps occur frequently, and that the "current" state of stocks cannot be appraised in those years when data are missing. As a rule, surveys should be annual to be usable safely in the deplorably polemical context of fisheries management.

Acknowledgements

This work was made possible with the financial contribution of the EC to the institutes participating in the FISBOAT project (research project DG-Fish, STREP n° 502572 (2004-2007)).

Dr Terry Quinn (University of Alaska, Juneau, USA) kindly provided the NRC simulated data sets.

References

Barrowman, N.J. and Myers, R.A. (2000). Still more spawner-recruitment curves: the hockey stick and its generalizations. *Can. J. Fish Aquat. Sci.*, **57**: 665-676.

Beare, D.J., Needle, C.L., Burns, F. and Reid, D.G. (2005). Using survey data independently from commercial data in stock assessment: An example using haddock in ICES Division VIa. *ICES Journal of Marine Science* **62**: 996–1005.

Beverton, R.J.H. and Holt, S.J. (1957). On the dynamics of exploited fish populations. U.K. Min. Agric. Fish., *Fish. Invest*. (Ser. 2), **19**: 533 pp.

Chen, S. and Watanabe, S. (1989). Age Dependence of Natural Mortality Coefficient in Fish Population Dynamics. *Nippon Suisan Gakkaishi*, **55**(2): 205-208.

Cook, R.M. (1997). Stock trends in six North Sea stocks as revealed by an analysis of research vessel surveys. *ICES Journal of Marine Science* **54**: 924–933.

Cook, R.M. (2004). Estimation of the age-specific rate of natural mortality for Shetland sandeels. *ICES Journal of Marine Science* **61**: 159–164.

Cotter, A.J.R. (2001). Intercalibration of North Sea International Bottom Trawl Surveys by fitting year-class curves. *ICES Journal of Marine Science* **58**: 622-632 [Erratum, Ibid. 58:1340].

Cotter, A.J.R. and Buckland, S.T. (2004). Using the EM algorithm to weight data sets of unknown precision when modeling fish stocks. *Mathematical Biosciences* **190**: 1-7.

Cotter, A.J.R., Mesnil, B. and Piet, G. (2007) Estimating stock parameters from trawl cpue-at-age series using year-class curves. *ICES Journal of Marine Science* **63**: 234-247.

Darby, C.D. and Flatman, S. (1994) Virtual population analysis: version 3.1 (windows/DOS) user guide. CEFAS, Lowestoft, UK.

Deriso, R.B., Quinn, T.J.II and Neal, P.R. (1985). Catch-age analysis with auxiliary information. *Canadian Journal of Fisheries and Aquatic Sciences* **42**: 815–824.

Fryer, R.J. (2002). TSA: is it the way? Appendix D in Report of Working Group on Methods of Fish Stock Assessment. ICES CM 2002/D:01.

Fournier, D. (2005). An introduction to AD MODEL BUILDER version 7.0.1 for use in nonlinear modeling and statistics. Available from http://otter-rsch.com/admodel.htm.

Gudmundsson, G. (1986). Statistical considerations in the analysis of catch-at-age observations. *Journal du Conseil International pour l'Exploration de la Mer* **43**: 83–90.

Gudmundsson, G. (1994). Time series analysis of catch-at-age observations. *Applied Statistics* **43**: 117-126.

Gulland, J.A. (1969). Manuel des méthodes d'évaluation des stocks d'animaux aquatiques. Première partie- Analyse des populations. *Manuels FAO de science halieutique*, N. 4 : 160 pp.

Hilborn, R. and Walters, C.J. (1992). *Quantitative fisheries stock assessment: Choice, dynamics and uncertainty*. Chapman and Hall, New York.

Johnson, S.J. and Quinn, T.J. II (1987). Length frequency analysis of sablefish in the Gulf of Alaska. *Technical Report UAJ-SFS-8714*, University of Alaska, School of Fisheries and Science, Juneau, Alaska. Contract report to Auke Bay National Laboratory.

Needle, C.L. (2002b). Preliminary analyses of survey indices for whiting in IV and VIId. Working Document WD2 to the ICES Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak, Copenhagen, June 2002.

Needle, C.L. (2002d). Survey-based assessments of whiting in VIa. Working Document WD1 to the ICES Working Group on the Assessment of Northern Shelf Demersal Stocks, Copenhagen, August–September 2002.

Needle, C.L. (2003d). Survey-based assessments with SURBA. Working Document to the ICES Working Group on Methods of Fish Stock Assessment, Copenhagen, 29 January – 5 February 2003.

Needle, C.L. (2004a). Absolute abundance estimates and other developments in SURBA. Working Document to the ICES Working Group on Methods of Fish Stock Assessment, IPIMAR, Lisbon 10–18 Feb 2004.

Needle, C.L. (2004b). Data simulation and testing of XSA, SURBA and TSA. Working Paper to the ICES Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak, Bergen, September 2004.

NRC, 1998. *Improving fish stock assessments*. National Academy Press, Washington, D.C., 177 p. (Appendix E describes the data generation; Appendix I shows plots of biomass trajectories).

Oehlert, G.W. (1992). A note on the delta method. *American Statistician* **46**: 27–29.

Patterson, K.R. and Melvin, G.D. (1996). Integrated Catch At Age Analysis Version 1:2. *Scottish Fisheries Research Report*. FRS: Aberdeen.

Pope, J.G. and Shepherd, J.G. (1982). A simple method for the consistent interpretation of catch-at-age data. *Journal du Conseil International pour l'Exploration de la Mer* **40**: 176–184.

Quinn, T.J.II and Deriso, R.B. (1999). *Quantitative Fish Dynamics*. Oxford University Press, Oxford.

Ricker, W.E. (1975). Computation and interpretation of biological statistics of fish population. *Bull. Fish. Res. Bd. Can.*, **191**: 382 pp.

Seber, G.A.F. (1982). *The Estimation of Animal Abundance*. 2nd edn, Griffin, London.

Shepherd, J.G. (1982). A versatile new stock-recruitment relationship for fisheries, and the construction of sustainable yield curves. *Journal du Conseil International pour l'Exploration de la Mer* **40** (1): 67-75.

Sparre, P. and Venema, S.C. (1998). Introduction to tropical fish stock assessment. Part 1, manual. *FAO Fish. Techn. Pap*., 306 (1) Rev. 2: 407 pp.

Thompson, W.F.and Bell, F.H. (1934). Biological statistics of the Pacific halibut fishery. 2. Effect of changes in intensity upon total yield and yield per unit of gear. *Rep. Int. Fish. (Pacific Halibut) Comm.* **8**: 49 p.

Trenkel, V.M. (2006). Combining acoustic and DEPM survey indices in the biomass random effects model for stock assessment. Working Document to the ICES Working Group on Acoustic and Egg Surveys for Sardine and Anchovy in ICES areas VIII and IX (WGACEGG), November 2006. 13 pp.

Trenkel, V.M. (2007). A biomass random effects model (BREM) for stock assessment using only survey data: application to Bay of Biscay anchovy. ICES CM 2007/O:03.

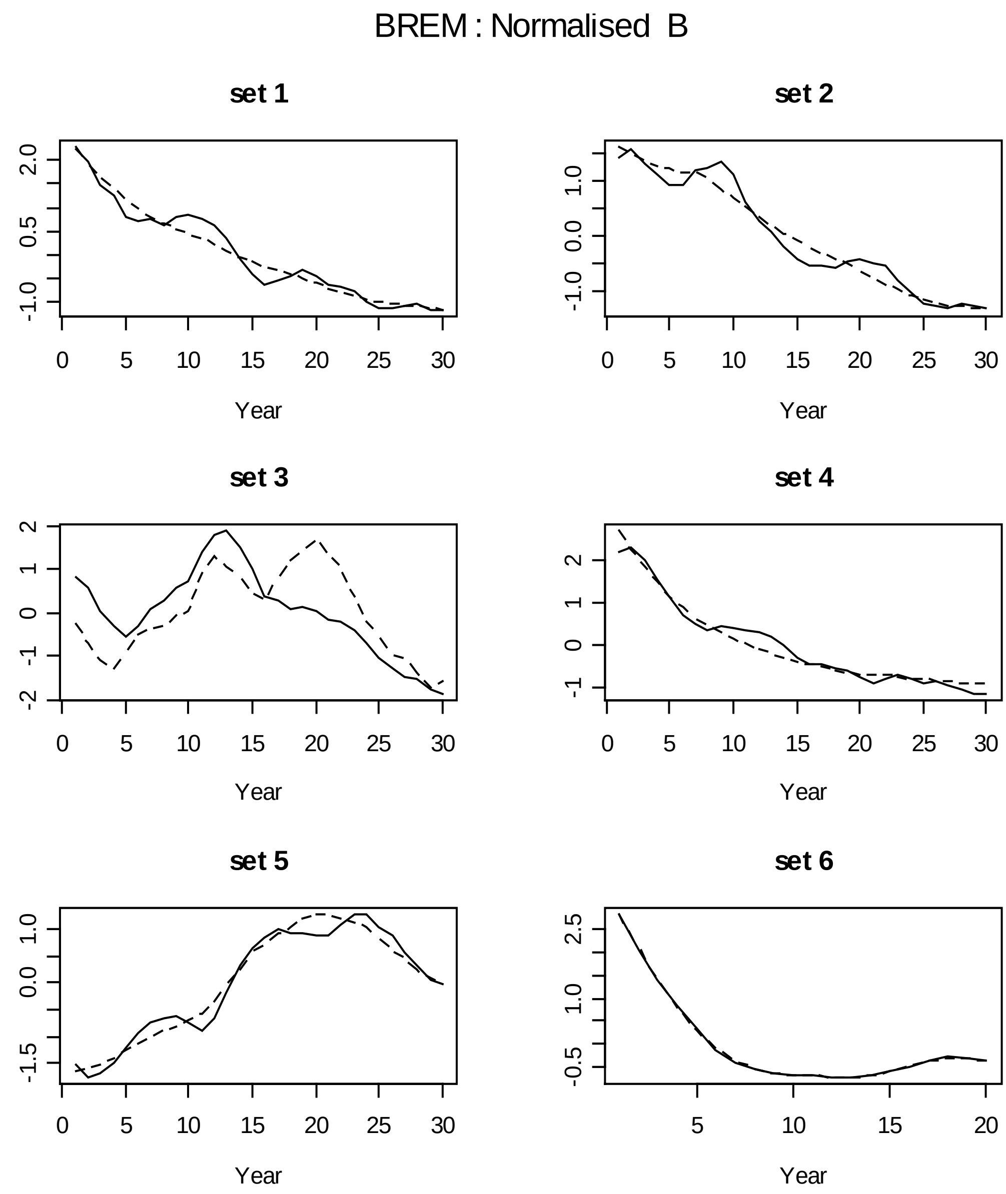

Fig. 1.a. BREM: Comparison of normalised series of biomass estimates (dashed) vs. truth (solid).

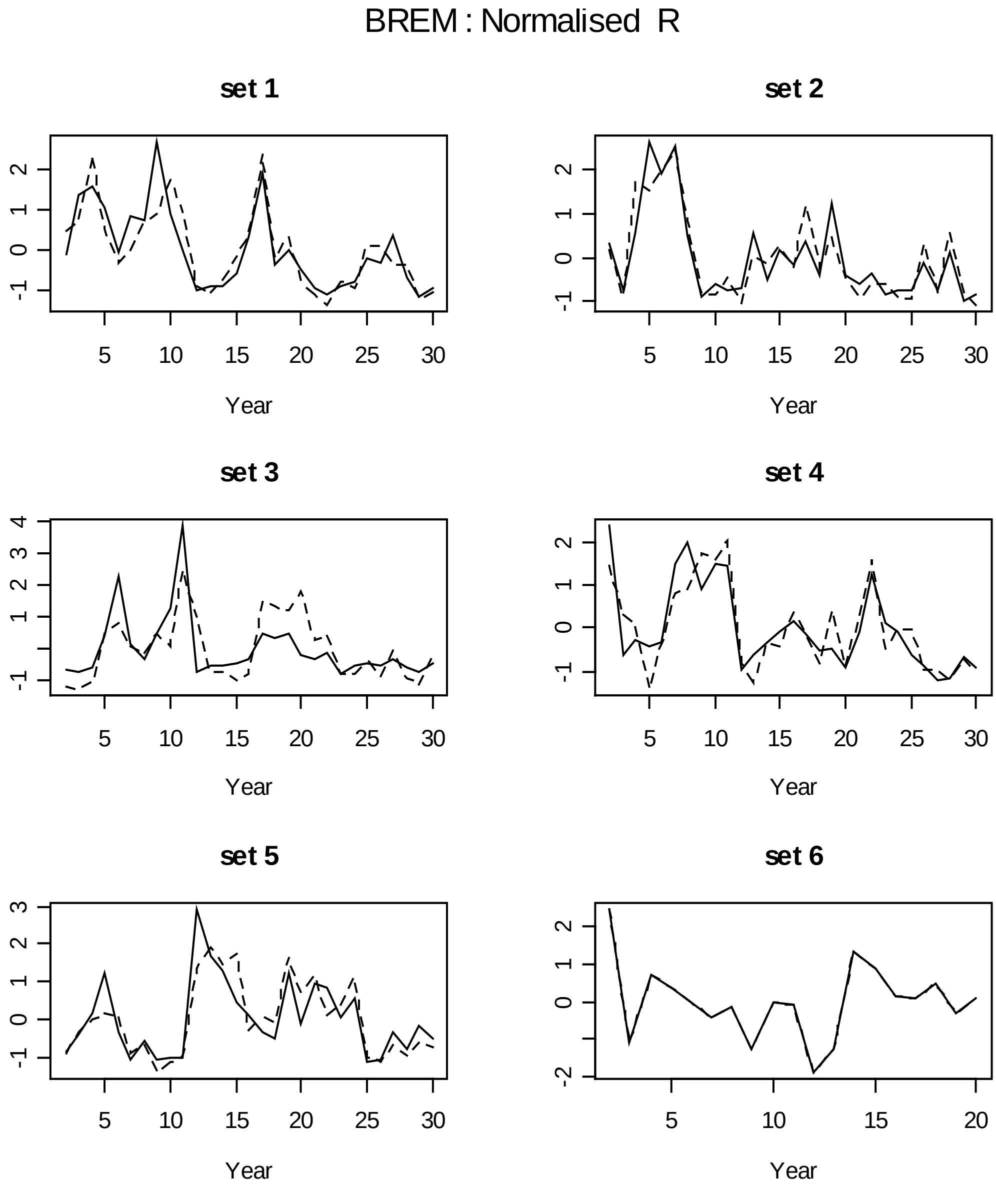

Fig. 1.b. BREM: Comparison of normalised series of recruitment estimates (dashed) vs. truth (solid).

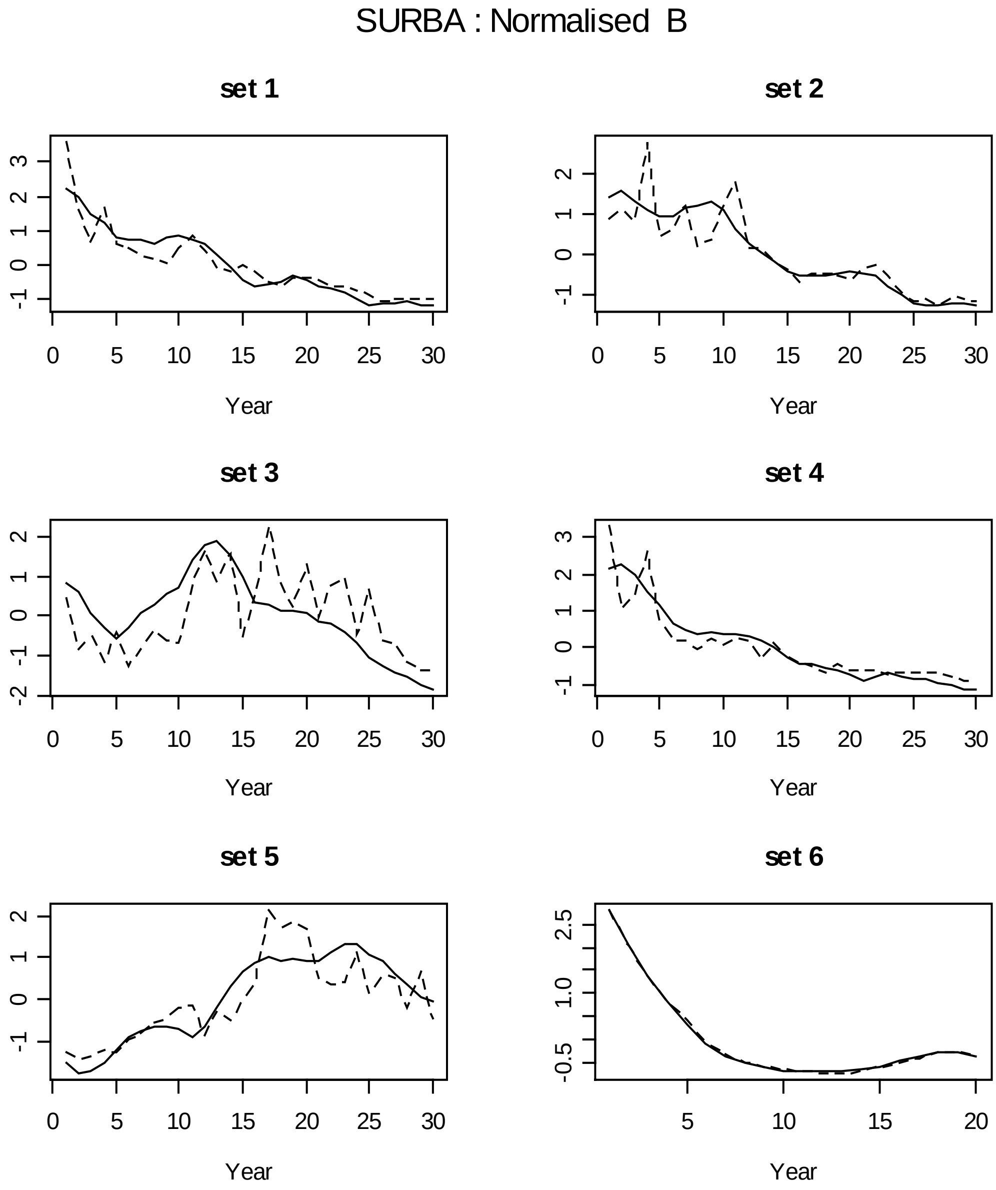

Fig. 2.a. SURBA: Comparison of normalised series of biomass estimates (dashed) vs. truth (solid).

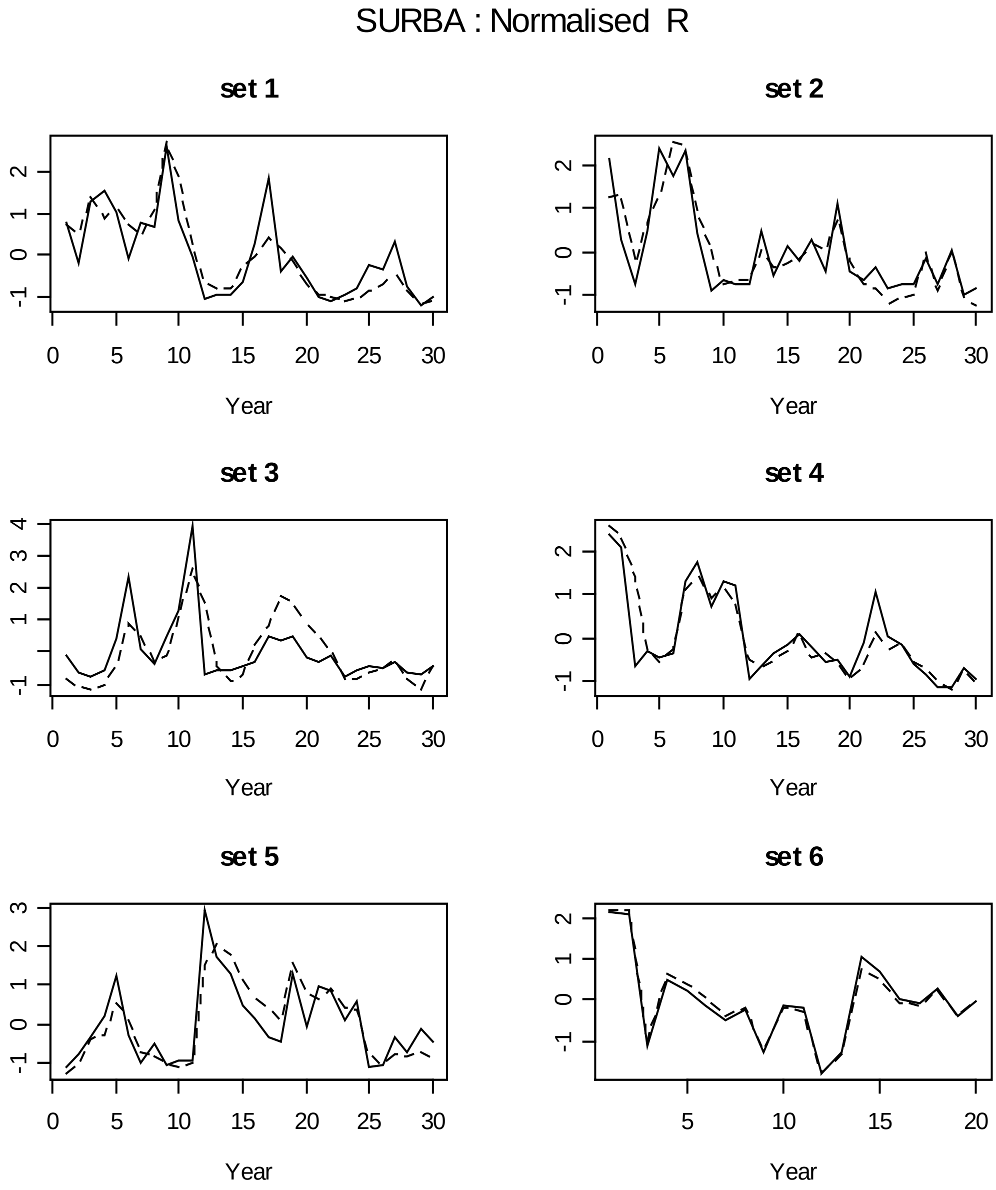

Fig. 2.b. SURBA: Comparison of normalised series of recruitment estimates (dashed) vs. truth (solid).

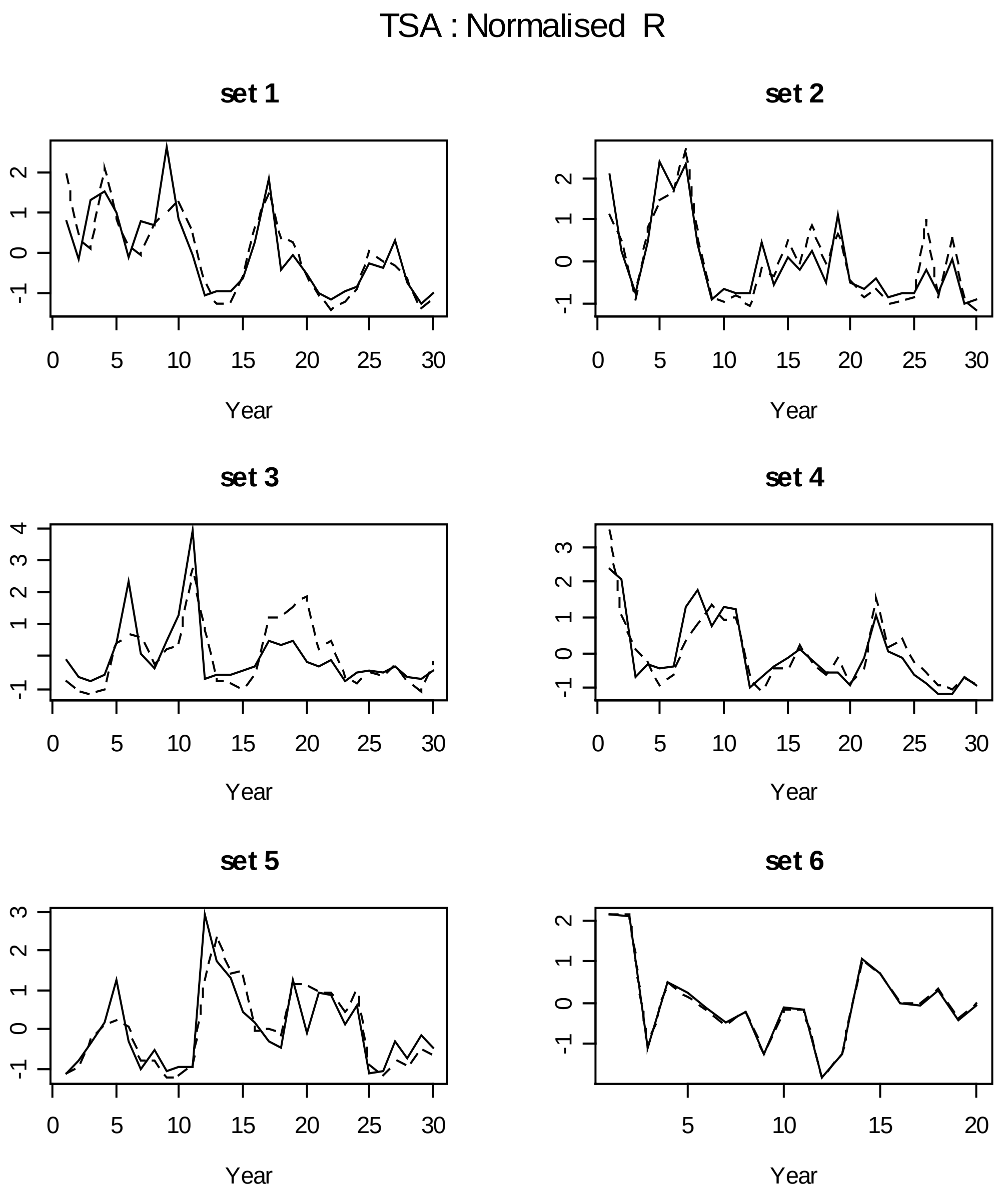

Fig. 3. TSA: Comparison of normalised series of recruitment estimates (dashed) vs. truth (solid).

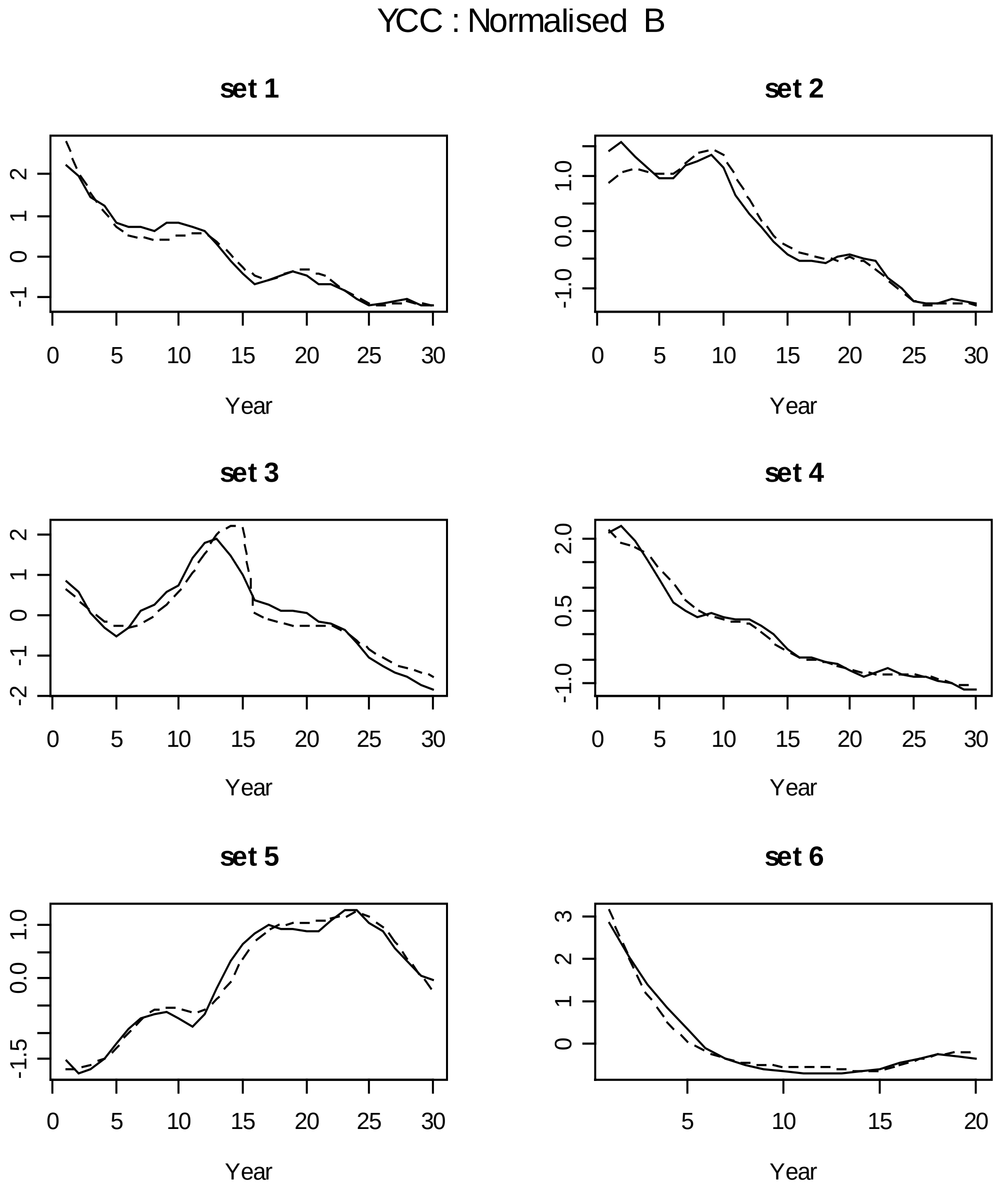

Fig. 4.a. YCC: Comparison of normalised series of biomass estimates (dashed) vs. truth (solid).
NB: set 3 = split survey (3.2)

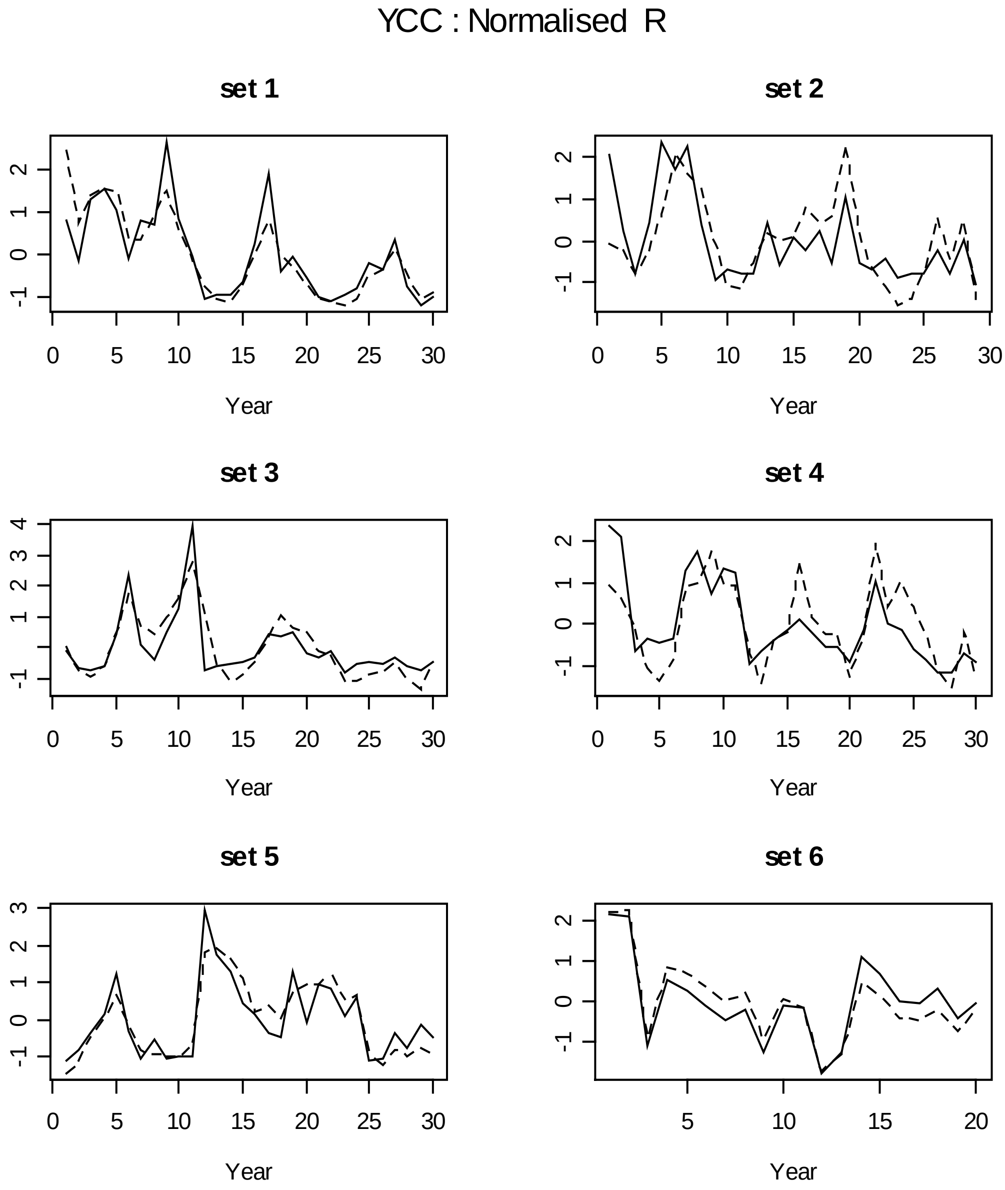

Fig. 4.b. YCC: Comparison of normalised series of recruitment estimates (dashed) vs. truth (solid).
NB: set 3 = split survey (3.2)