ICES SGRF REPORT 2012

ICES ADVISORY COMMITTEE

ICES CM 2012/ACOM:24

Report of the Study Group on Recruitment Forecasting (SGRF)

15–19 October 2012

Barcelona, Spain

ICES International Council for the Exploration of the Sea
CIEM Conseil International pour l'Exploration de la Mer

International Council for the Exploration of the Sea
Conseil International pour l'Exploration de la Mer

H. C. Andersens Boulevard 44–46
DK-1553 Copenhagen V
Denmark
Telephone (+45) 33 38 67 00
Telefax (+45) 33 93 42 15
www.ices.dk
info@ices.dk

Recommended format for purposes of citation:

ICES. 2012. Report of the Study Group on Recruitment Forecasting (SGRF), 15–19 October 2012, Barcelona, Spain. ICES CM 2012/ACOM:24. 40 pp.
https://doi.org/10.17895/ices.pub.19281179

Contents

1 Executive summary

The Study Group on Recruitment Forecasting (SGRF) met at the Institute of Marine Sciences (ICM-CSIC) in Barcelona, Spain from October 15–19, 2012, with eleven participants and Dr Sam Subbey (Norway) as Chair.

The formal mandate for this SG meeting was established in 2011/2/ACOM26 under Action Plan No: 1.2, 1.10, and 2.5. The objectives of the SG are to decide on guidelines and standards with regards to (1) How to develop models for recruitment projections which incorporate both abundance indices and environmental drivers, and (2) Criteria for validating models and for choosing the 'best' or a set of the best models.

The 2012 meeting focused on:

a) A framework for:
   - i) Ecological and biological considerations in developing models for short-term recruitment projections, including how to incorporate both abundance indices and environmental drivers;
   - ii) The detection of ecological drivers for spikes in fish recruitment;
   - iii) A methodological approach to deriving a representative recruitment forecast when confronted with an ensemble of several competing model forecasts of recruitment;

b) Testing the framework using a designated case study.

This report summarizes work by the SG in devising practical ecological and biological guidelines for developing short-term recruitment forecasting models. The report also discusses how to deal with multiple model recruitment forecasts and presents an illustrative example using recruitment models for Northeast Arctic Cod (NEAc).

The SG recognizes that understanding the ecological underpinning to recruitment is a prerequisite in the development of appropriate models, which link fish recruitment to biotic and abiotic process drivers. A good understanding of such drivers is necessary e.g. for predicting spikes in fish recruitment.

The SG advocates combining recruitment forecasts from several candidate models, rather than forecasts from an individual 'best' model. The variance across a number of models is related to the risk of selecting among these models. Hence the goal of combining individual forecasts will be to reduce the variance of the performance across the combinations relative to the variance across the individual methods, for various measures of variance. A guiding principle is that irrespective of the methodology adopted in combining forecasts, the predictive performance of the combination must be better than that of the selected individual models.

Throughout the rest of this report, short-term forecasts will refer to forecasts one year ahead. With respect to stock assessment however, the expression 'short-term forecast' usually refers to predictions associated with year classes that have already been spawned but yet to enter the fishery (See ICES 2011 SGRF for a discussion).

Significantly absent from the suite of recruitment forecasting models for the designated case study are the classical Beverton–Holt and Ricker functions. The SG investigated the performance of these classical functions using data from the case study. Since these classical functions do not include explicite terms for environmental drivers, the results are presented separately, in the appendix.

2 A framework for recruitment modelling; Incorporating abundance indices and environmental drivers

2.1 Ecological considerations

Recruitment is the result of many factors, starting with the parents, which affect survival from the egg-stage through to when individuals recruit to the fishery or stock. One can treat recruitment as a stepwise process (egg, larvae, juvenile, adult), where abundance at one stage is a function of abundance at a previous stage (Paulik, 1973; Rothschild, 1986; Figure 1). The state of a population in any given year is a function of the stock (e.g. reproduction, growth, biomass) and recruitment, which is itself a function of past events (e.g. state of the stock, environmental conditions). Recruitment is therefore explicitly linked not only to the amount of spawning–stock biomass (SSB), but also parental size and growth history of the individual. Recruitment must decline if there is insufficient spawning biomass, but may also decline when the growth environment of either the prerecruit or the post-recruit (but sexually immature) portion of the stock changes (Lorenzen, 2008; Lorenzen and Enberg, 2002). Reduced body size will result in decreased recruitment because fecundity, egg size, and spawning extent is inextricably linked to the individual's growth history and condition (Kjesbu *et al*., 1996). Recruitment predictions that make incorrect assumptions regarding the spawning stock, ignore non-linear feedback mechanisms, or omit interactions between stages and across spatio-temporal scales tend to lead to a breakdown of the stock–recruitment relationship (Hutchings and Rangeley, 2011; Myers, 1998; Neill *et al*., 1994).

The stock–recruitment curve captures the transition from recruitment, which integrates the previous effects of these factors into the dynamics of a single cohort, back into the spawning–stock biomass (Skjoldal, 2004). Whether variation in life-history traits might affect the stock–recruitment relationship or if these relationships can be captured by certain factors should be determined. Diagramming multistage stock–recruitment relationships (see Figure 1) are an important and visually easy-to-use aid for assessing which life-stage might be important to investigate. The eventual reproductive success of an individual is determined by the conditions the individual experienced up to the point of maturation; any change to the environment of the individual will subsequently conclude in change in recruitment. These changes are described by the growth dynamics of an individual; once recruitment is viewed from a bioenergetics prospective, it is apparent that growth at the prerecruit and immature, post-recruit stages will affect recruitment (see Enberg *et al*., 2011; van Der Veer *et al*., 1994). Individuals utilize an optimal temperature range in order to maximize energy intake, the difference between the potential and realized growth is referred to as the scope-for-growth (Jobling, 1994; Neill *et al*., 1994). Even if under suboptimal conditions, the individual will attempt to maximize the scope between growth and maintenance costs (Pörtner and Farrell, 2008). Individuals typically do not achieve their full potential for growth under natural conditions (Figure 2; Dutil and Brander, 2003); however, growth should increase when conditions improve, even if some conditions are still considered suboptimal. Under declining population size, conditions for individual growth are expected to increase as a result of decreased density-dependent effects (Rose, 2005). However, when population depletion is not accompanied by increased growth (e.g. many Canadian cod populations), per capita population growth rate (r) is predicted to decline (Hutchings, 2005).

Biological and physical environmental factors tend to influence recruitment directly (e.g. mortality of eggs and larvae) or indirectly, through changes in predation, prey availability, growth, and, thus, parental trait plasticity. Furthermore, any change that results in the contraction of the spatial extent of habitat or productivity of the system can lead to a decline in recruitment; it is not until the system shifts states (or the spatial extent of habitat expands) that the species will then begin to show stronger recruitment (Alvarez-Fernandez *et al*., 2012). This effect is not directly related to stock size, as corroborated by the variable recruitment at similar spawning stock sizes. Additional changes are attributed to the effect of fisheries and selective harvesting (Enberg *et al*., 2010). Fishing affects recruitment indirectly, primarily through spawner biomass and traits, and directly, should the fishery include sexually immature fish (Vasilakopoulos *et al*., 2011). The relationship between stock and recruitment will shift with demographic changes and any changes in fecundity, size and age-at-maturity, or spatio-temporal spawning extent of the population. Changes in size and age structure of the population will invariably lead to changes in the production of viable offspring (Marteinsdottir and Begg, 2002; Trippel, 1995); the effect is more chronic in short-lived species. Furthermore, if spawning–stock biomass is a factor included in the model, recruitment estimates must consider if there is a portion of the stock that does not spawn every year (Skjaeraasen *et al*., 2012). Due to the potential phenomenon of 'skipped spawning' the maturity ogive and SSB needs to be clearly defined (see ICES 2012, WKMATCH). SSB can be considered as all fish which are sexually mature (a sexually mature fish is defined by WKMATCH as: the individual has the capability to enter, either regularly or continuously, the gonadotropin-dependent reproductive cycle with the resulting production of sex steroids and activation of related hormonal receptors with the proviso that once a fish is sexually mature, it remains that way for the rest of its life). Therefore, care must be taken to ensure all sexually mature fish are included in the maturity ogive. When considering the relationship with recruitment, then factors such as skipped spawners need to be taken into account and the SSB should be adjusted to remove those individuals that will omit a spawning and not contribute to the egg production or viable offspring.

Recruitment projections

Short- and medium-term projections of recruitment are needed for stock assessments. These projections should be based upon generation time of the species under review (see (ICES 2011, SGRF). Projections in the short- and medium-term for short-lived species will therefore be at a shorter time-scale than long-lived species.

2.1.1 Spiked recruitment

Sporadic, exceedingly strong (or poor) recruitment pulses are often referred to as spiked recruitment, which can be defined as periods of exceptionally high (or low) survival for the early life stages (Figure 3). Over longer time periods (i.e. centuries), spikes may become more apparent, whereas short time-scales (i.e. decades) may merely show an increasing (decreasing) trend in the stock. Here, we will only consider strong, relatively short-term pulses. Norwegian spring-spawning herring is often presented as a good illustration of this phenomenon (Figure 4). Spiked recruitment requires a certain level of biomass, which differs by stock, and low biomass does not rule out the possibility of such dynamics; North Sea herring has shown spiked recruitment at very low biomass. As long as the stock is not in the density-independent phase of stock–recruitment, a spike may occur. Moreover,

spiked recruitment can appear across many species at the same time (e.g. plaice and cod stocks in a number of different adjacent management areas in the Northeast Atlantic; Fox *et al*., 2000).

Spiked recruitment is an indication that something has happened within the system outwith the effects of the spawning stock, providing an indication that conditions should be closely monitored. Temperature conditions have been linked to these pulses for species that inhabit water near the edge of their thermal optimums (O'Brien *et al*., 2000; Planque and Frédou, 1999), but this is not the sole factor driving such dynamics (Petitgas *et al*., 2011). Shifts in productivity, measured through multiple parameters, often underlie such outbursts (Munch and Kottas, 2009); shifts in productivity can be viewed by the relationship between SSB and recruitment, split into different environmental situations/regimes (Figure 4; Olsen *et al*., 2011). When conditions align spatial-temporally (e.g. temperature, match with prey, low cannibalism), there is an appropriate response in recruitment. This aligning of conditions can be thought of as an elevator that raises the population to a new state (Figure 5; Solari *et al*., 1997). Under these circumstances, the stock–recruitment relationship can be complex. Furthermore, dynamics within the stock are often driven for years afterwards by these strong pulses (Skjoldal, 2004), particularly when they result in strong density-dependent responses, where recruitment of subsequent year classes (at the appropriate lag) is depressed (Caley *et al*., 1996) or when they change the behaviour of the stock (e.g. Huse *et al*., 2010).

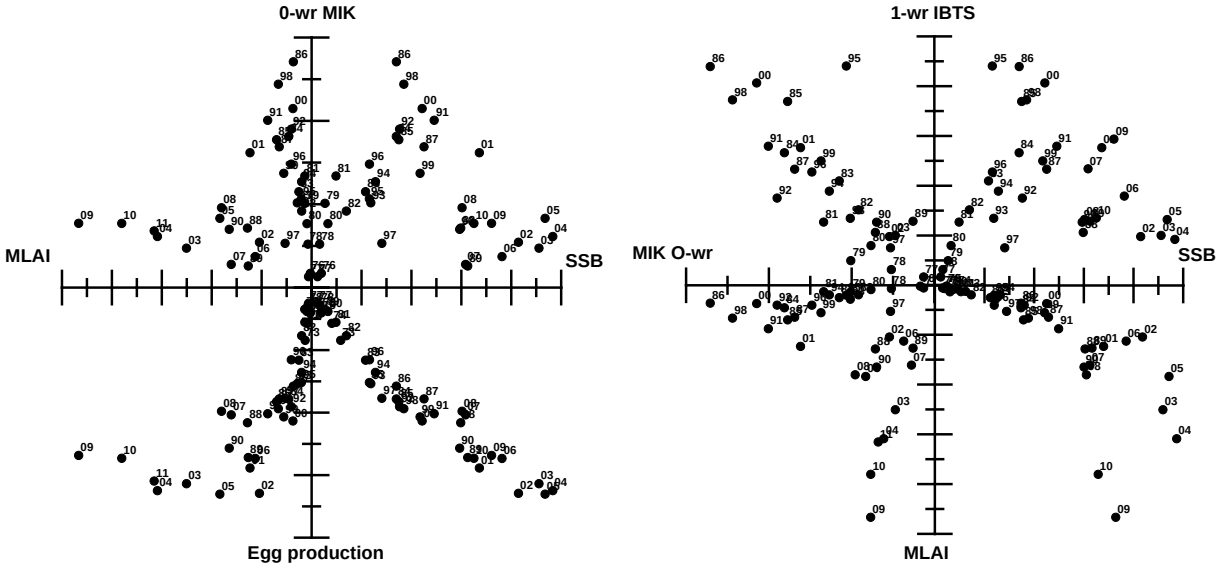

Figure 1. A multistage stock–recruitment relationship for North Sea autumn herring (updated from Payne *et al*., 2009). (Left) Relationship between spawning stock – egg-production – larval abundance (MLAI) – 0-group. (Right) Relationship between spawning stock – larval abundance (MLAI) – 0-group – age-1 group. See Nash and Dickey-Collas (2005) for details on data sources and methodologies.

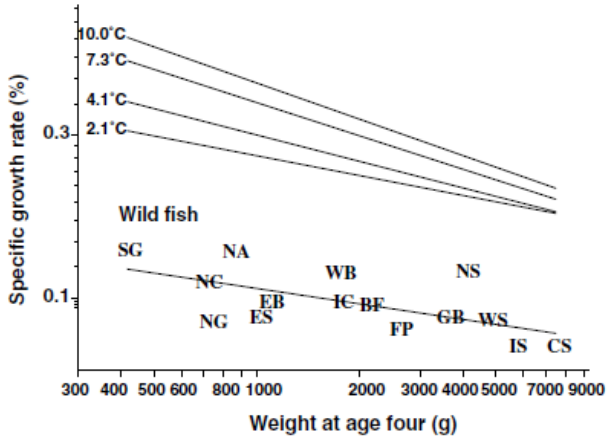

Figure 2. Food unlimited change in growth rate for Atlantic cod from age 4 to age 5 at several different temperatures (upper lines) compared to change in growth rate for 15 stocks of Atlantic cod (lower line); taken from Dutil and Brander (2003).

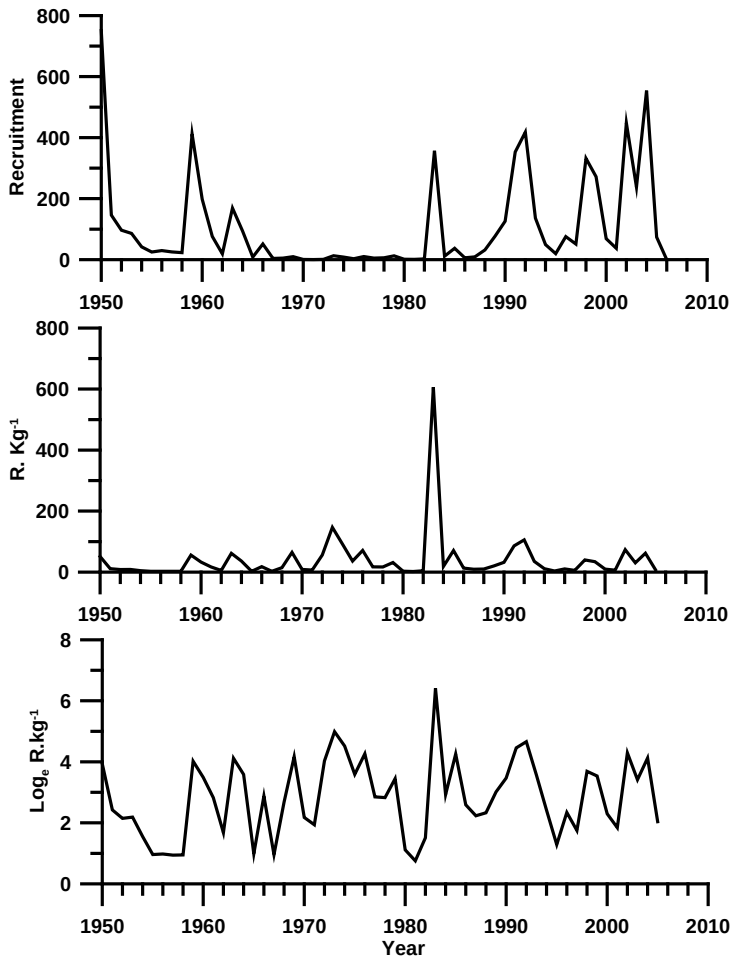

Figure 3. Top panel: Recruitment over time for Norwegian spring-spawning herring, showing periodic peaks in recruitment. Mid panel: Index of survival rate (R/kg SSB) over time indicating peak in survival for one of the strong recruitment pulses. Lower panel: Survival index on the log scale shows a large amount of variability exists in this index over time and indicates that even during a low period of stock size, survival rate was approximately average (1970–1980).

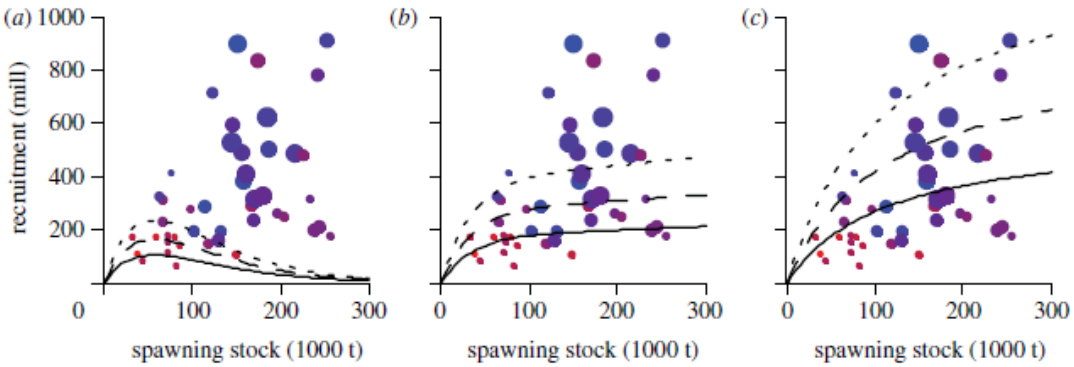

Figure 4. Effect of different environmental conditions on the stock–recruitment relationship for North Sea cod; taken from Olsen *et al*. (2011).

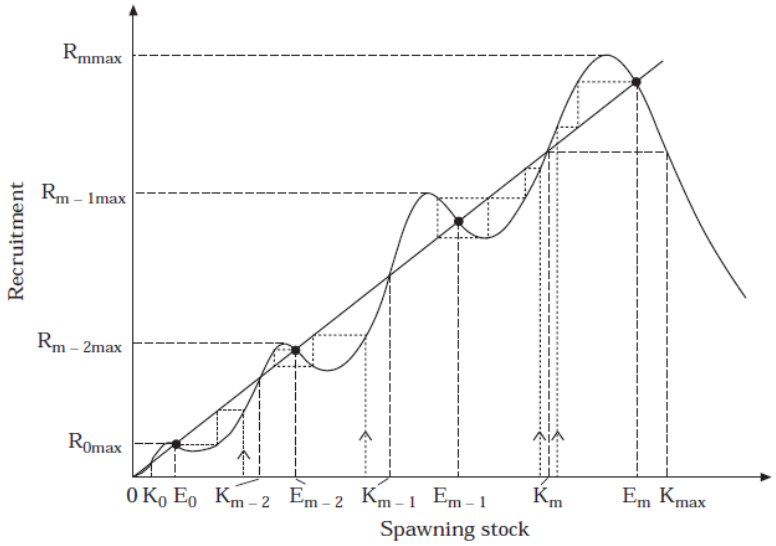

Figure 5. Multiple stable states proposed for the Baltic cod stock–recruitment relationship, 1973–1993, taken from Solari *et al*. (1997).

3 Criteria for single model evaluation

Diagnostic tools are required to evaluate model fit to observations, fidelity to underlying assumptions, and to assess model predictions of recruitment. Standard hypothesis testing methods (F-tests, likelihood ratio test, score test, etc.) can be used to compare complicated (highly parameterized) to less complicated models (McCullagh and Nelder, 1989). Such methods are however, only directly applicable to nested models. Information-theoretic methods, e.g. Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) are applicable to non-nested models. A brief description of information-theoretic methods is provided below.

3.1 AIC(c)

If we define k as the number of parameters in the statistical model and L as the maximized value of the likelihood function for the estimated model, then the AIC is defined as:

$AIC=2k-2ln(L)$ or alternatively, $AIC=\chi 2 + 2k$.

The latter form is often convenient, because most model-fitting programs produce $\chi 2$ as a statistic for the fit. Given a set of candidate models for the data, the preferred model is the one with the minimum AIC value. Hence AIC not only rewards goodness-of-fit, but also includes a penalty that is an increasing function of the number of estimated parameters, and thus limits over-fitting. The classical AIC, though, is only valid asymptotically. Thus if the number of datapoints is small (such as in recruitment data), then the corrected AIC, AICC must be used. The AICc is defined as:

$AICc = AIC + 2k(k+1)/(n-k-1)$,

Where n is the number of datapoints Burnham and Anderson (2002) strongly recommend using AICc, rather than AIC, if n is small or k is large. Since AICc converges to AIC as n gets large, AICc generally should be employed regardless, (Burnham and Andersen, 2002).

3.2 BIC

A convenient formulation of the Bayesian Information Criterion (BIC) is given by

$BIC= \chi^2 + k\, ln(n).$

A comparison of AIC/AICc and BIC is given by Burnham and Anderson (2002, Section 6.4). The authors argue that AIC/AICc has theoretical advantages over BIC. The authors present a few simulation studies that suggest AICc tends to have practical/performance advantages over BIC. In particular, AIC is asymptotically optimal in selecting the model with the least mean squared error, under the assumption that the exact "true" model is not in the candidate set (as is virtually always the case in practice); BIC is not asymptotically optimal.

3.3 Cross-validation

Cross validation involves using a 'training dataset' (a subset of the total dataset) to estimate the parameters of the model and using the resulting model to predict the remaining data (the 'test dataset'). The ability of the model to predict the test set is used to select the explanatory variables to include in the model. If too many explanatory variables are used, one obtains good model fit for the training data but a model with poor predictive ability since the good fit also includes fitting noise, rather than

only signal. If too few explanatory variables are used, the model performance is poor for both test and training datasets. K-fold validation is a version of cross validation that makes more use of the data than simple cross.-validation. In K-fold CV, the data are divided into k-equal parts and the model is run k-times, each time rotating through each of the k-subsets as the test dataset and using the remaining data as the training dataset. A potential drawback with cross validation is that it does not necessarily parallel the likelihood inference that is used to estimate the model parameters. This is because a test criterion is required, and simple least squares are often used (Trevor Hastie *et al*., 2001). However, the likelihood function may differ from the least-squares criterion.

3.4 Dealing with multiple recruitment model predictions

In general, the task of choosing the 'best' model among a variety of candidates is a statistically challenging and non-trivial problem. For a review, see de Gooijer, Abraham, Gould and Robinson (1985). When short time-series are used as input, it is hard to distinguish between closely related models (based e.g. on AIC, BIC) since selection indices tend to be very close to each other. A change in for instance, the length of the input data, may result in a different model choice, and consequently in the forecast. See a detailed discussion in Zou, H. and Yang, Y. (2004). When candidate models use different datasets or different combinations of such in the modelling process, choosing a 'best' candidate model becomes an even more challenging task especially when the variable to be forecasted is inherently highly uncertain. This can happen when the variable to be forecasted is derived for instance, from models which take uncertain data as input, e.g. forecasting recruitment based on model derived estimates of spawning–stock biomass. Thus choosing a 'best' recruitment model based on comparing forecasts with e.g. recruitment values from an assessment model could be misleading and risky. Put in other words, when there is uncertainty about the 'best' individual forecasts or combination, it might be riskier to select among individual forecasts than to select among their combinations, Hibon, M. and Evgeniou, T. (2005).

Combining individual model forecasts as introduced by Bates and Granger (1969) is often considered as a successful alternative to using just an individual forecasting method. Further, there is theoretically proven advantage of a proper combining over any selection method, see Yang (2004). Specifically for time-series forecasting, it has been shown that predictive performance increases through combining forecasts (Armstrong, 1989; 2001; Clemen, 1989; Makridakis and Winkler, 1983).

This SG therefore advocates for combining recruitment forecasts from several candidate models, rather than forecasts from an individual 'best' model. The variance across a number of models is related to the risk of selecting among these models (Bousquet and Elisseeff, 2002; Vapnik, 1998). Hence the goal of combining individual forecasts will be to reduce the variance of the performance across the combinations relative to the variance across the individual methods, for various measures of variance (Breiman, 1998; Evgeniou, Pontil and Elisseeff, 2004). Irrespective of the methodology adopted in combining forecasts, the combined forecasts must be such that (i) the predictive performance of the combination is better than that of the selected individual models and (ii) the risk, measured as the difference in post-sample performance between the selected and the best possible, for selecting among individual methods is no higher than the risk for selecting among combinations.

The literature contains several methodologies for dealing with multiple competing models and model predictions. This report presents an example methodology, which

generates a representative prediction through a weighted average of multiple predictions. The weights are determined by the historical performance of each model, and updated as new data are included in the analysis.

3.4.1 Example methodology; the AFTER algorithm

The SGRF meeting in 2012 has tested an example approach which involves combining weighting individual forecasts. The approach is based on using the *Aggregated Forecast Through Exponential Reweighting* (**AFTER**) algorithm. Individual model weights are based on past model prediction performance. A brief description is provided below. However, a detailed methodological description can be found in Yang, Y. (2004) and Zou, H. and Yang, Y. (2004).

3.4.1.1 Algorithm description and assumptions

It is assumed that

1) The conditional distribution of Y_i given Y_{i-1} is Gaussian for all $i \geq 1$ with conditional mean, m_i, and conditional variance, v_i;

2) For each forecasting procedure, in addition to the forecast, $\hat{y}_n$, an estimate of $\hat{v}_n$ is obtained based on Y_{i-1}.

To combine forecasts from J models, at each time n, the AFTER algorithm looks at their past performances and assigns weights accordingly as follows. Let Wj,1 =1/J and for $n \geq 2$, let

$$W_{j,n} = \frac{\Pi_{i=1}^{n-1} \frac{1}{\sqrt{\hat{v}_{j,i}}} \exp\{-\frac{1}{2}\sum_{i=1}^{n-1} \frac{(Y_i-\hat{y}_{j,i})^2}{\hat{v}_{j,i}}\}}{\sum_{k=1}^{J} \Pi_{i=1}^{n-1} \frac{1}{\sqrt{\hat{v}_{k,i}}} \exp\{-\frac{1}{2}\sum_{i=1}^{n-1} \frac{(Y_i-\hat{y}_{k,i})^2}{\hat{v}_{k,i}}\}} \quad (1)$$

Then the combined forecast, $\hat{y}_n^{\dagger}$ is given by

$$\hat{y}_n^{\dagger} = \sum_{j=1}^{J} W_{j,n}\hat{y}_{j,n} \quad (2)$$

Note that after each additional observation, the weights on the candidate forecasts are updated, and that the weight *Wj,n* depends only on the past forecasts and the past realizations of Y.

This report illustrates the application of the above weighted averaging methodology using recruitment models for northeast Arctic cod. A brief description of the models is provided in the next section, as well as details about model runs and analysis.

4 Designated case study

The SG used the northeast Arctic cod (NEAc) as a primary case study. Though a description of this stock is provided in the SGRF 2011 report, this is repeated here for the sake of completeness. The NEAc stock spawns along the Norwegian coast, mainly north of 67°N, during March-April. Most of the larvae drift into the Barents Sea, where the cod spends the rest of its life, except for the spawning migration. Recruitment to the fisheries has been at age three. The NEAc is an opportunistic feeder, foraging mostly on available species of suitable size, though capelin appears to be its preferable prey. In years when capelin abundance is low in relation to the cod density, cannibalism may cause a substantial mortality on juveniles. Other important predators of cod are seals and whales.

4.1 Model descriptions, simulation results and diagnostics

Since 2008 several regression models, which includes stock and climate variables, have been used by the Arctic Fisheries Working Group (AFWG) for prediction of NEA cod recruitment-at-age 3 (ICES CM 2011/ACOM: 05). Although several recruitment models exist only four of such models have been chosen to demonstrate how competing model predictions can be combined into generating a representative statistic with reduced variance. Detailed description of the models can be found in the SGRF 2011 report.

The calibration dataseries is from the period 1984 to 2008. The calibrated model was then used for recruitment predictions in the period 2009–2010. For the retrospective analysis, the models were run, excluding data one year at a time. The shortest time-series used was from 1984 to 1999 and the longest was for the period 1984–2008 (full series). The 2012 AFWG VPA assessment was used as "truth".

4.1.1 Model 1—JES

This model was first described in Stiansen *et al*. (2005), and further evaluated in Subbey *et al*. (2008). The model is given as:

JES: $R3_t \sim Tw_{t-3} + Age1_{t-2} + \log(CapMatBio_{t-2})$,

where Tw the water temperature: 3–7 stations of the Kola section (layer 0–200 m), Age1 is the bottom-trawl abundance (age 1 year index) of NEA cod from the joint winter Barents Sea acoustic survey and CapMatBio is the biomass of mature capelin in October. The numbers in parentheses are the time-lags in years.

Table 1 is a table of correlation coefficients between the covariates of the model, and shows the highest correlation between the estimates of recruitment and Tw.

Table 1. Correlation of model time-series.

	R3 (VPA)	Tw	Age1	CapMatBio
R3(VPA)	1	0.69	0.29	0.27
Tw	0.69	1	0.01	0.24
Age1	0.29	0.01	1	-0.2
CapMatBio	0.27	0.24	-0.2	1

4.1.1.1 Retrospective analysis and discussion

The model gave a good fit to the VPA until around 2006, from when the fit (R2) started to decrease. From 2007 the capelin term became insignificant, and the same applied to the Age 1 term from 2008.

This may serve as an example of a model that does no longer captures the dominant dynamics influencing the stock recruitment. Thus, recruitment prognoses from this model cannot be considered as single accurate estimates. However, in line with the approach adopted in this report, recruitment forecasts from JES may be included in defining averaged ensemble prognosis of stock recruitment. The prognosis given by the model is 924 072 (for 2009) and 952 637 (for 2010), in thousand individuals.

The model mean square prediction error is estimated through retrospective analysis. Table 2 shows a summary of the retrospective analysis and the estimated mean square prediction error (MSE), which is calculated as the average of the squared difference between the VPA and the prognoses.

Table 2. Calculation of mean square error (MSE), based on the retrospective runs.

	VPA	P1 (Prognosis one year ahead)	P2 (Prognosis two years ahead)	diff (vpa–p1)^2	diff(vpa–p2)^2
2000	613 588	577 427.4		1 307 588 992	
2001	520 652	530 838	524 701.8	103 754 596	16 400 880.04
2002	454 916	688 360	689 438.4	54 496 101 136	55 000 756 102
2003	709 786	811 730.9	843 472	10 392 762 636	17 871 946 596
2004	310 760	677 055.3	692 105.5	1.34172E+11	1.45424E+11
2005	580 528	713 233.3	734 483.5	17 610 696 648	23 702 295 980
2006	602 424	456 457.1	463 067	21 306 335 896	19 420 373 449
2007	1 345 611	601 367.7	584 910.6	5.53898E+11	5.78665E+11
2008	1 180 149	823 126.8	681 847.2	1.27465E+11	2.48305E+11
MSE				1.02306E+11	1.36051E+11
stdev				319 853	368 851

4.1.2 Model 2—RCT3

Similar to last AFWG (ICES, 2008) the following default settings have been chosen for running RCT3 during the SGRF 2012 meeting:

1) Regression type = C;
2) Tapered time weighting applied;
3) Power = 3 over 20 years;
4) Survey weighting not applied;
5) Final estimates shrunk towards mean;
6) Minimum S.E. for any survey taken as 0.20;
7) Minimum of 3 points used for regression;
8) Forecast/Hindcast variance correction used.

The input data for the model were time-series survey data for ages 0, 1 and 2 from the Russian autumn survey and for ages 1, 2 and 3 from the joint Norwegian-Russian winter survey. There are two types of indices available from the joint winter survey; acoustic and bottom-trawl estimates. In contrast to the AFWG runs where both types of data were used, only acoustic indices have been used here, since they show closer relationship with recruitment-at-age 3 from the VPA. The list of the six chosen indices is as follows:

1) R-0 Russian Swept-area trawl survey, area I+IIb, age 0
2) R-1 Russian Swept-area trawl survey, area I+IIb, age 1
3) R-2 Russian Swept-area trawl survey, area I+IIb, age 2
4) N-BSA1 Norwegian Barents Sea Acoustic survey age 1
5) N-BSA2 Norwegian Barents Sea Acoustic survey age 2
6) N-BSA3 Norwegian Barents Sea Acoustic survey age 3

It was observed that predictions for year classes 2007 and 2008 are highly driven by the shrinkage parameter, and the mean VPA estimates used in shrinkage diverge considerably from the survey predictions. It should also be mentioned that the standard errors for some indices are extremely high. An alternative model configuration without shrinkage has been tested in retrospective analysis (see below).

The prognosis given by the model is 529 (for 2009), 390 (for 2010) and 690 (for 2010) in million individuals. The results are summarized in Table 3.

Table 3. RCT model predictions and associated variance estimates of recruitment.

YEAR	VPA 2012 (MILL. IND.)	PROGNOSIS FOR NUMBER OF YEARS AHEAD OF 2008			ASSOCIATED VARIANCES		
		1 year	2 years	3 years	1 year	2 years	3 years
2009	750	529			109 541		
2010	457		390			121 680	
2011	691			690			123 945

4.1.2.1 Retrospective analysis and discussion

With the exception of two year classes (2004 and 2005), Figure 6 shows that the model gives a reasonably good prediction (compared to the VPA) of NEA cod recruitment-at-age 3.

A possible explanation for the less accurate prediction of year classes 2004 and 2005 could be the shrinkage procedure used in RCT3 program, as this limits the model ability to predict big changes in recruitment. An extra run has been performed to test the influence of shrinkage on model performance. RCT3 runs without shrinkage demonstrate slightly better predictions for year classes with outstanding abundance. However, the quality of predictions for other year classes, become less accurate, Figure 7.

Another possible explanation for model inability to predict 2004 and 2005 year-class strength is a possible error in the VPA estimates. These two points clearly deviate from the general pattern in the VPA/index regressions, see Figure 8.

Comparison of the last three years assessment values shows that their estimates are rather unstable and belongs to unconverted part of VPA (Figure 9). As it was recognized by AFWG (ICES 2012) the current VPA estimates become very sensitive to XSA model assumption regarding density-dependent survey catchability. Hence a possible error in these year class estimates may explain the failure of all recruitment models to predict them correctly.

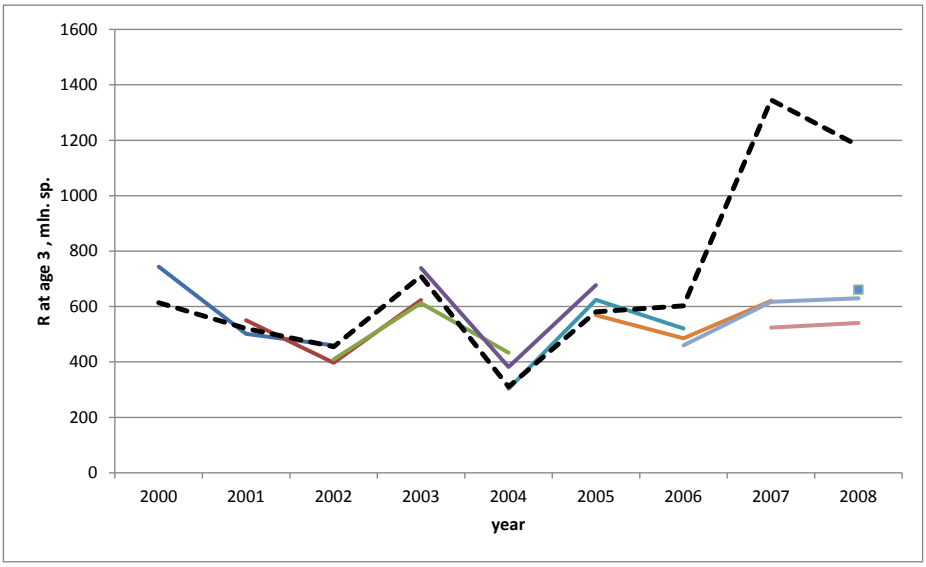

Figure 6. Recruitment forecasts for 1, 2 and 3 years (coloured lines), using the RCT3 model. The figure shows a comparison to the VPA estimates of NEA cod recruitment-at-age 3 (black dotted line). The mean square deviations of predicted age 3 recruitment values by the RCT3 model were in range of 47–49% of the average recruitment for the period (see Table 4).

Table 4. Absolute and relative mean square deviations of predicted RCT3 recruitment values from VPA estimates for northeast Arctic cod at age 3.

	PREDICTIONS		
	1 year	2 years	3 years
Mean square deviation (RMSE)	331	349	352
N (number of years in prediction)	9	8	7
Average Recruitment (mill. Ind.)	702	713	741
Mean percentage deviation in	47	49	48

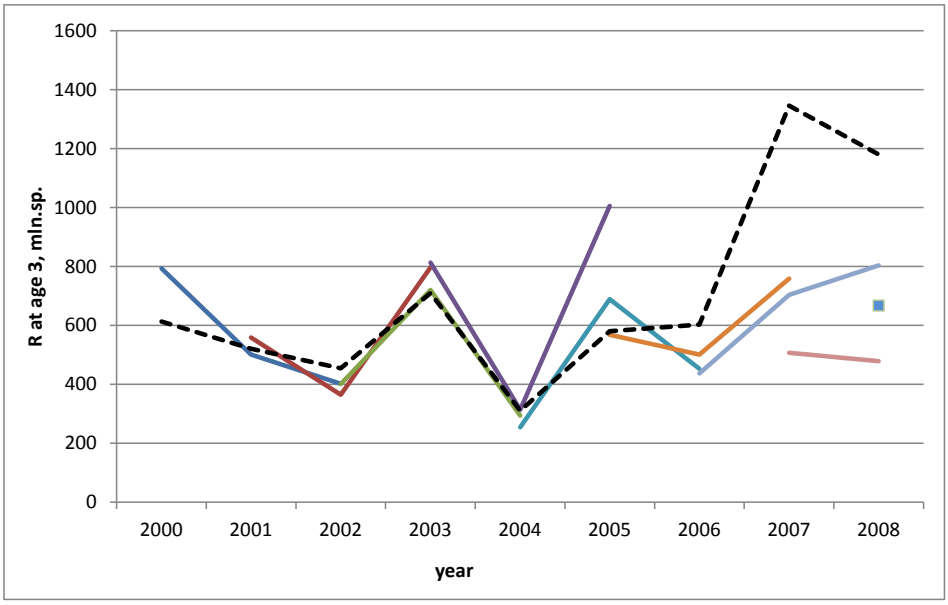

Figure 7. Results of 1, 2 and 3 years' predictions (coloured lines) by RCT3 in comparison to estimates of NEA cod recruitment-at-age 3 from VPA (black dotted line). The results are based on no shrinkage to the mean recruitment.

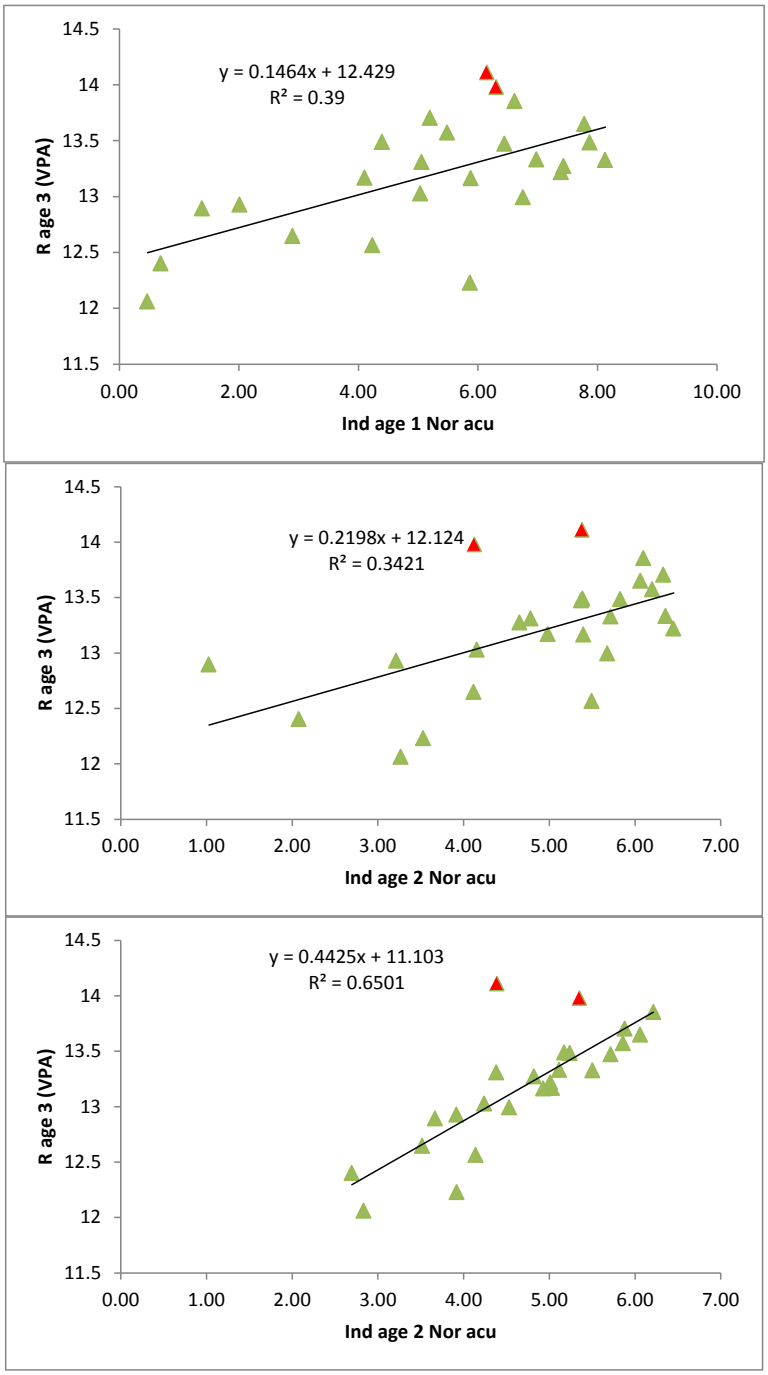

Figure 8. Regressions between VPA recruitment-at-age 3 and acoustic indices from the joint winter survey at ages 1–3. The axes are on logarithmic scale. The 2004 and 2005 year classes are denoted by red triangles.

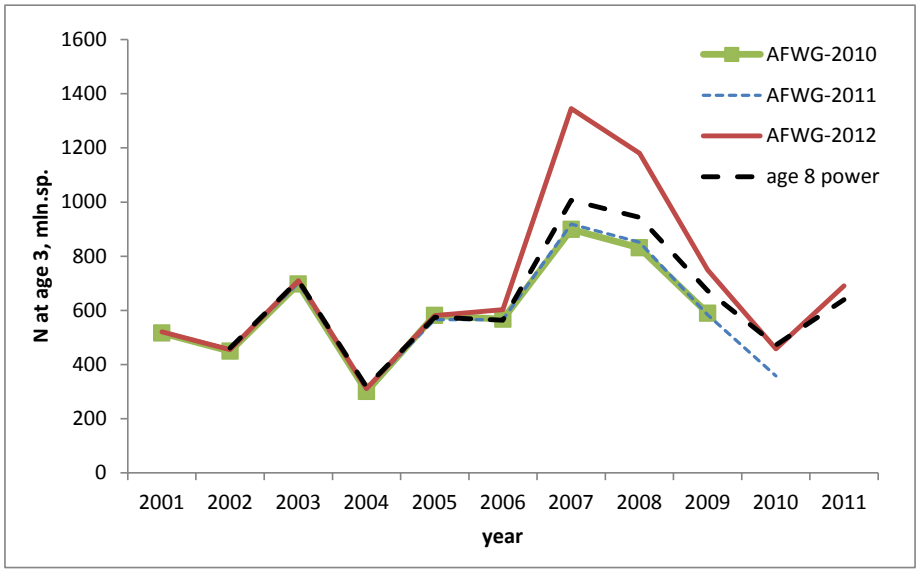

Figure 9. Estimates of NEA cod recruitment-at-age 3 by the AFWG (2010–2012) and recruitment estimate using the 2012 assessment data, which assumes a power relationship between VPA numbers and survey index at age 8 (black dotted line).

4.1.3 Model 3 and 4—TITOV

An approach to model stock–recruitment by including indices which reflect variations in the physical and chemical environment has been implemented for the Barents Sea capelin and northeast Arctic cod (Titov, 1999; Titov, 2001). The statistical models were revised in 2009 (Titov *et al*., 2005; Titov, 2008). In the revision, data prior to, and including the year 1983 were excluded in the model calibration and analysis in 2010. In order to improve prediction, water temperature data were added as an explanatory variable in one of the models in 2011. Some terms were also dropped from the model formulation in order to avoid over-fitting the regression models, and thus improve model prediction power. This was done in 2011 in accordance with statistical criteria (Titov, 2011).

A description of the input data and two of the models (Titov3 and Titov2) used in the case study for the SGRF 2012 meeting is presented below.

4.1.3.1 Data and model description

1) (Ta) mean monthly anomalies of air temperature at the Murmansk station;

2) (Tw) mean monthly anomalies of water temperature at stations 3–7 of the Kola section (0–200 m layer);

3) (I) mean monthly anomalies of ice coverage of the Barents Sea (percentage ratio between the area covered by ice and total area);

4) (OxSat) mean monthly anomalies of saturation by oxygen of near-bottom-water layers at 3–7 stations of the Kola;

5) (Cod3) values of abundance of cod at the age of 3 (Anon, 2012);

6) (CodC0) values of 0-group cod abundance index(Anon, 2012);

7) (CodA1) sum of acoustic abundance of cod at the age of 1 (Anon, 2012);

8) (CodA2) sum of acoustic abundance of cod at the age of 2 (Anon, 2012);

9) (CodA3) sum of acoustic abundance of cod at the age of 3 (Anon, 2012);

10) (SSB) values of spawning part biomass of cod population (Anon, 2012).

Indices ITa and DOxSat are calculated on basis of I, Ta and OxSat (Titov, 2011).

4.1.3.2 Linking recruitment (Cod3) with abiotic and biotic parameters

The final set of predictors was determined by the method of step-by-step multiple regression, using the Statgraphics Plus package for Windows 2.1. The equation for the forecast of Cod3, 1 year (Titov1) and 2 years (Titov2) in advance based on parameters estimated in 2011 are given below.

(Titov2) $Cod3_t \sim DOxSat_{t-13}$ ^2+ $ITa_{t-39} + CodA1_{t-23} + Tw_{t-17}$

(Titov3) $Cod3_t \sim ITa_{t-39} + \log(CodC0_{t-28}) + Tw_{t-26}$

Both statistical models had p-values <0.01, corresponding to 99% level of significance.

For this particular report, each model was tuned to the time-series from 1984 to 1999. Independent forecasts were made for the period 2000–2011. Values of the standard deviation for each of the models (RMSE) are calculated on basis of data for 2000–2008. Prognoses from models (Titov3 and Titov2) are shown in Table 5.

Table 5. Recruitment models prognoses using Titov 2.

YEAR	VPA 2012 (MILL. IND.)	PROGNOSIS FOR NUMBER OF YEARS AHEAD OF 2008			ASSOCIATED VARIANCES		
		1 year	2 years	3 years	1 year	2 years	3 years
2009	750	727			240		
2010	457		561			272	
2011	691						

Table 6. Recruitment models prognoses using Titov 3.

YEAR	VPA 2012 (MILL. IND.)	PROGNOSIS FOR NUMBER OF YEARS AHEAD OF 2008			ASSOCIATED VARIANCES		
		1 year	2 years	3 years	1 year	2 years	3 years
2009	750	621			242		
2010	457		490			271	
2011	691			381			293

4.2 Multiple recruitment predictions; ensemble averaging

To illustrate the application of the weighted averaging we used results from the set of four models in the case study. The models were calibrated using data for the period 1994–2008. The AFTER algorithm was then applied to model predictions (with estimates of prediction variance) of recruitment for 2009 and 2010. Table 7 shows a tabulated comparison of the predicted values while Figure 10 shows the graphical representation. Observe that in general, the variance of the averaged model (when compared to the VPA estimates in retrospect) is lower than variance range defined by the complete ensemble of models, see Figure 10.

Table 7. Individual model predictions of recruitment-at-age 3 for northeast Arctic cod.

YEAR	JES	RCT3	TITOV2	TITOV3	AFTER	VPA
2009	924	529	727	621	700	750
2010	953	390	562	490	591	457

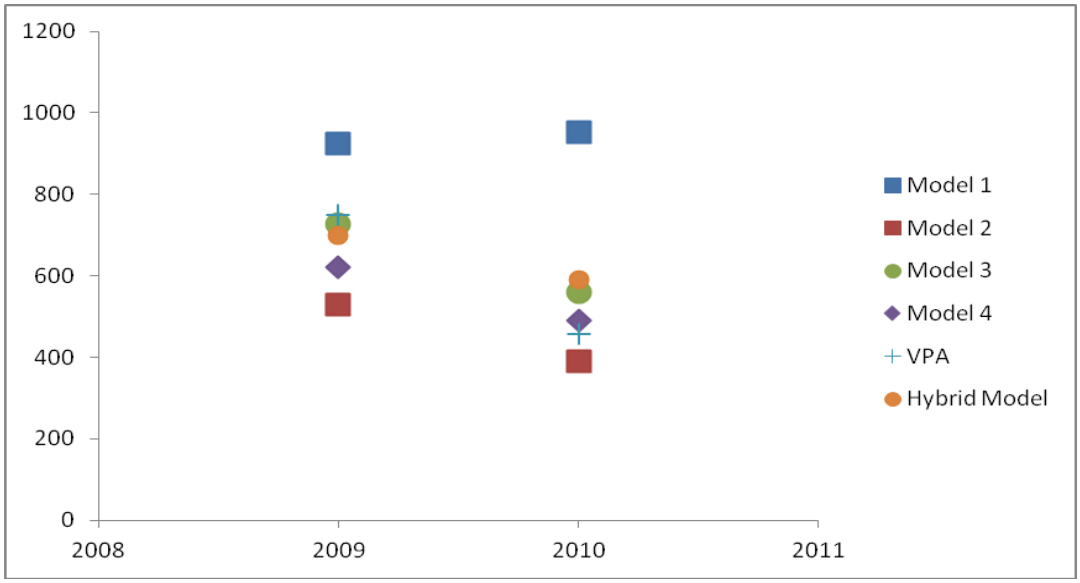

Figure 10. Graphical rendering of results in Table 7.

5 Conclusions and further directions

Both the biotic and abiotic environment can influence the survival of early life-history stages and the stock–recruitment drivers are most probably attributable to multiple factors. Since the ecosystem is subject to dynamic fluctuations, the principal drivers, or combination of drivers, may not always be the same. Models for recruitment forecasting which do not include environmental drivers may be limited in their ability to capture and predict e.g. spikes in stock recruitment. However, there must be a clear biological/ecological reasoning for the inclusion of one or several environmental time-series as drivers for recruitment. Further, when there is time-lag between time-series representing environmental drivers and recruitment, such time-lag must have ecological/biological underpinning and supported by sound statistical evaluations.

Since the input data (time-series) are invariably uncertain, quantifying risk and uncertainty must be important components of recruitment forecasting. The combined effect of model choice being influenced by several factors (choice of data type, scale, length and condition), and uncertainties (both in the data and models) implies that there is no single, correct forecasting model. A hybrid forecast based on combining recruitment forecasts from several candidate models, rather than forecasts from an individual 'best' model is a conservative approach, which reduces the risk of forecast failure. The literature addresses several methods for combining model forecasts. This report presents an example approach. Irrespective of the methodology adopted in combining forecasts however, the combined forecasts must possess lower variance and be such that the predictive performance of the combination is better than that of the selected individual models.

The results presented in this report only address issues related to short-term recruitment forecasts. The meeting in 2013 will aim at extending the analysis and case studies to forecasts in the medium term, as well as include other stocks.

6 Recommendations

The SG suggests meeting in Barcelona in xx–xx October 2013. For the meeting in 2013, the group has identified the need to apply the framework developed in 2011–2012 to the following stocks:

1) North Sea cod (NS cod);

2) Norwegian spring-spawning herring (NSS herring);

3) North Sea Autumn Spawning herring (NSAS herring).

There is therefore a need to actively involve stock–recruitment models and modellers working with these stocks. The group also suggests an extension of the meeting duration from four days (at present) to at least twice as long (eight days), to allow for adequate time to address the 2013 ToRs and case studies.

Appendix A: Forecasting Barents Sea cod recruitment using simple stock–recruit models with autocorrelated errors

Methodology

The method only requires a time-series of stock–recruit estimates, $(R_1,S_1),\ldots,(R_T,S_T)$. It is similar to the approach of Needle *et al*. (2003), although these authors adjusted for autocorrelation in residuals after fitting the stock–recruit models, whereas in the current approach this adjustment is incorporated as part of the stock–recruit model fitting. The approach is very similar to Minto (2011).

Recruitment is considered to be random and is modelled as a simple function of stock size. For simplicity stock size is not considered to be random. Let μ(S) denote a model that gives the value of R at some level of S. Two simple parametric models are considered:

1) Beverton–Holt (BH): $\mu(s) = \alpha s/(\beta+s)$.
2) Ricker (RK): $\mu(s) = \alpha s\exp(-\beta s)$.

The statistical estimation model is $LOG(R_T) = LOG\{\mu(S_T)\} + \varepsilon_{ot} + \varepsilon_{pt}$, where ε_o is independent and identically distributed (iid) observation error and εp is μ process error that is assumed to be autocorrelated. Both these errors are assumed to be normally distributed. The observations errors, $\varepsilon_{ot} \sim N(0, \sigma_o^2)$, are not of interest but the process errors are. They represent departures in recruitment from the simple model μ. The process errors are assumed to be AR(1) autocorrelated,

$$\varepsilon_{pt} = \varphi\varepsilon_{pt-1} + \delta_t$$

Where

$$\delta_t \sim N\left\{0, \frac{\sigma_p^2}{1-\varphi^2}\right\}, t = 2, \ldots$$

are iid. This variance assumption for δ means that the stationary variance for ε_{pt} is σ_p^2. Also, $\varepsilon_1 \sim N\{0, \sigma_p^2\}$. AD Model Builder (ADMB Project 2009) with the random effects module was used to implement the model.

A stock–recruit model for Barents Sea cod can be used to forecast recruitment (at age three) in the next three years based on the current estimate of SSB and the estimates for the last two years.

The model was applied to data obtained from the most recent assessment. SSB estimates for 2010–2011 are not yet converged and were not used in the stock–recruit model.

Retrospective analyses were used to assess forecast accuracy; however, retrospective stock–recruit data were not available. Only the time-series from the most recent assessment was available. To mimic the effect of VPA variability of recruitment estimates, the three most recent recruitment values were discarded when performing retrospective forecasts. No retrospective assessment error was added to the SSB estimates but this should not affect recruitment estimates for the assessment year (year 0 projection; P0) because Barents Sea cod recruit at age 3 and current year recruitment is derived from SSB of three years ago which should be converged in the VPA. This also applies to one year forecasts (i.e. P1). Forecast or prediction error was measured

using the root mean squared difference between retrospective predictions and the 2011 VPA results, which are taken to be the best estimates available and used as a baseline.

Therefore, in this application the stock–recruit model is used to predict recruitment for the three years with omitted data, and to also predict recruitment in the next three years (i.e. the forecast). In total there are six predictions.

Results

Beverton–Holt

The measurement error variance estimate was at a lower bound (0.007). All of the variability was accounted for by process errors ($\hat{\sigma}_p$ = 0.59). Minto (2011) implemented essentially the same model for this stock and found a very similar result. The negative log-likelihood was 43.71. The stock–recruit data and model fit, ignoring process errors, are shown in Figure 1. While highly variable, there is evidence of a relationship between stock size and recruitment.

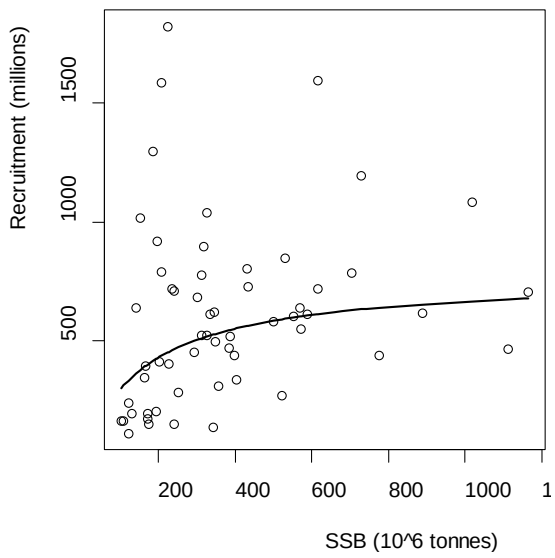

Figure 1. VPA estimates of SSB and subsequent recruitment-at-age 3 for Barents Sea cod. The solid line is a fitted BH curve.

Really poor recruitment has only occurred at low stock sizes, but good recruitment is possible over a wide range of stock size. Diagnostics of the model fit (Figure 2) look reasonable; however, it is clear that residuals are autocorrelated. Temporally adjacent residuals are usually similar in value. There is also no evidence of serious model misspecification in Figure 3.

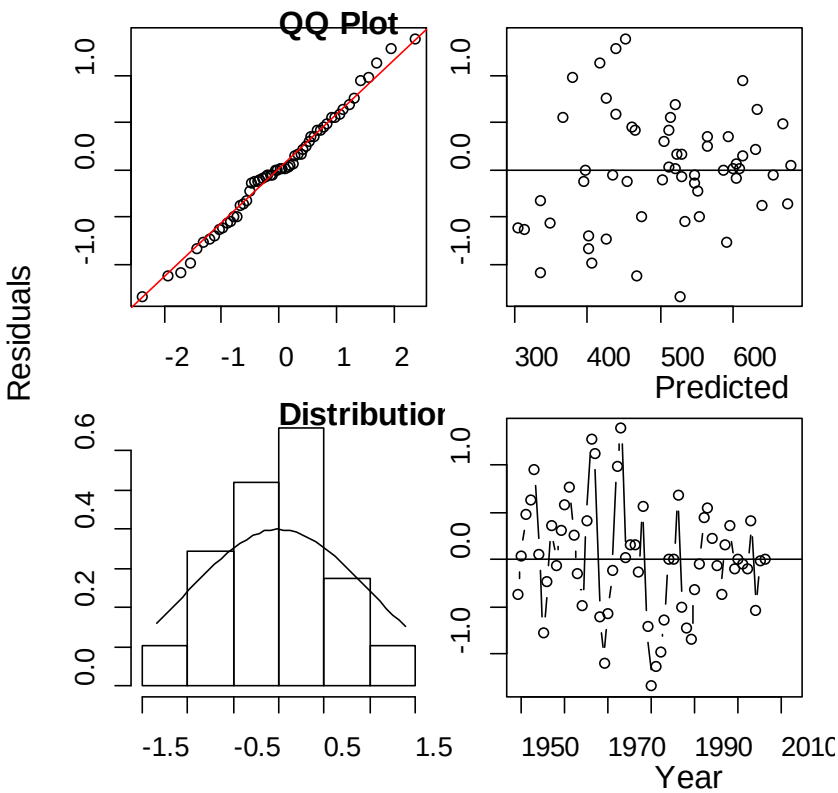

Figure 2. Residual diagnostic plots.

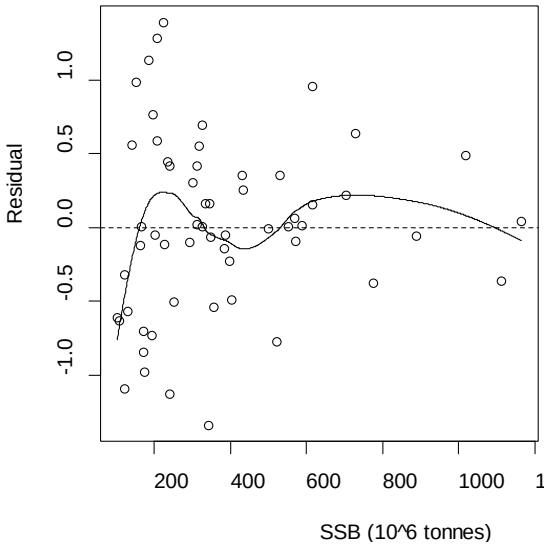

Figure 3. Residuals vs. stock size. The solid line indicates a loess (R, 2011) smooth of the residuals.

The model interpolated recruitment. Predictions of unobserved recruitments are shown in Figure 4 (triangles). Predictions of the size of the 2004–2005 year classes were quite different from the VPA estimates, but these estimates were contained in the 95% prediction intervals.

A comparison of retrospective predictions and current VPA estimates of recruitment is shown in Figure 5. Differences between forecasts and VPA results are also shown

in Figure 6. The root mean squared errors of these differences (for each projection period) are shown in Table 1. The values are large and this is caused by the anomalous VPA estimates of the 2004–2005 year classes. None of the retrospective predictions of the size of these year classes came close to the 2011 VPA values. The retrospective errors increased slightly with the length of the forecast, which is expected, although the increase is small.

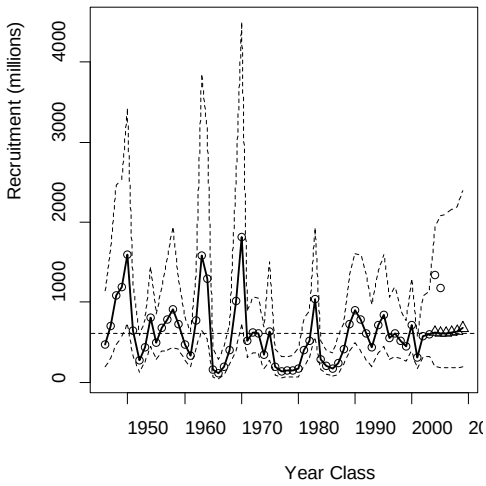

Figure 4. Model estimates (solid lines) and predictions (triangles) of recruitment for Barents Sea cod. The dashed lines indicate 95% prediction intervals. Circles are the converged VPA estimates, which are considered to be the best available estimates and used as a baseline.

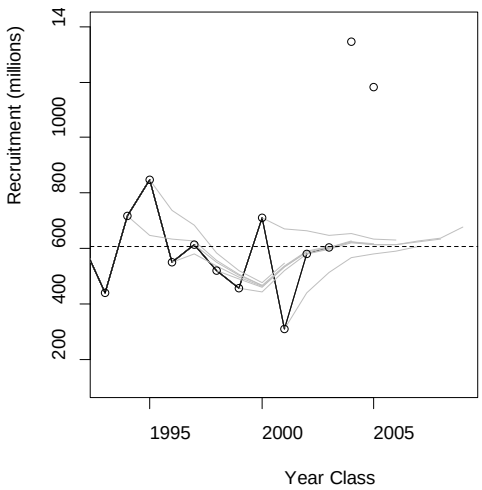

Figure 5. Retrospective estimates (solid lines) and predictions (grey lines), including three year forecasts, of recruitment. The model interpolated recruitment estimates, and the retrospective solid lines all overlap. Circles are the converged VPA estimates, which are considered to be the best available estimates and used as a baseline. The dashed line represents the average of the VPA estimates for 1946–2008; the time frame considered to be converged in the current assessment.

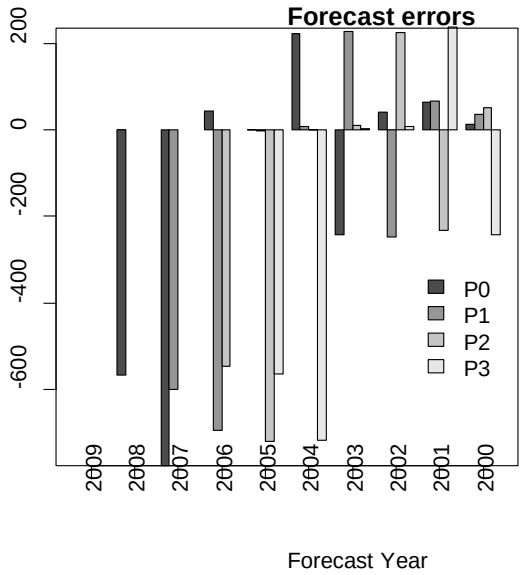

Figure 6. Differences between 2011 VPA estimates of recruitment to 2008 and retrospective stock–recruit model predictions for the current year (P0) and three year forecasts (P1–P3). Some P0–P4 differences could not be computed for 2006–2008 because some of the forecasts in these years were for recruitments later than 2008.

Table 1. Root mean squared error of retrospective predictions for the current assessment year (P0) and three year forecasts (P1–P3).

MODEL	P0	P1	P2	P3
Beverton–Holt	340.3	346.0	363.3	397.2
Ricker	340.2	346.1	362.4	377.1

Ricker

The measurement error variance estimate was also at a lower bound (0.007); however, not all retrospective estimates hit the bound but they were all close. All of the variability was accounted for by process errors ($\hat{\sigma}_P$ = 0.59). The negative log-likelihood was 43.75 which was approximately the same as the BH model fit. The stock–recruit data and the model fit, ignoring process errors, are shown in Figure 7.

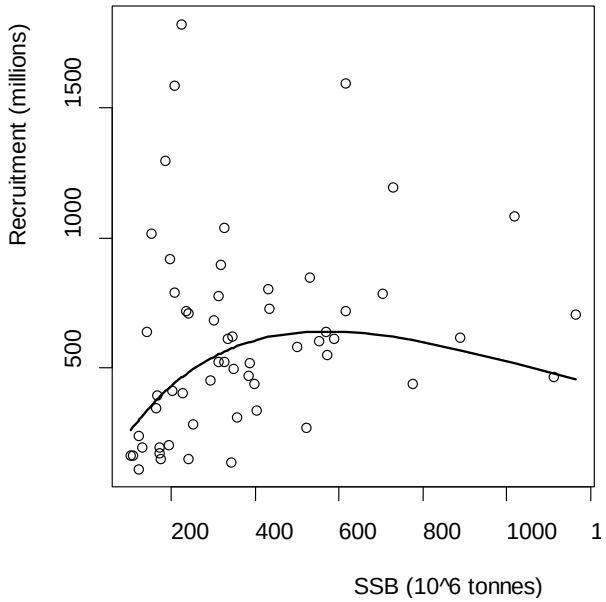

Figure 7. VPA estimates of SSB and subsequent recruitment-at-age 3 for Barents Sea cod. The solid line is a fitted Ricker curve.

Diagnostics of the model fit (Figures 8 and 9) look similar to the BH results.

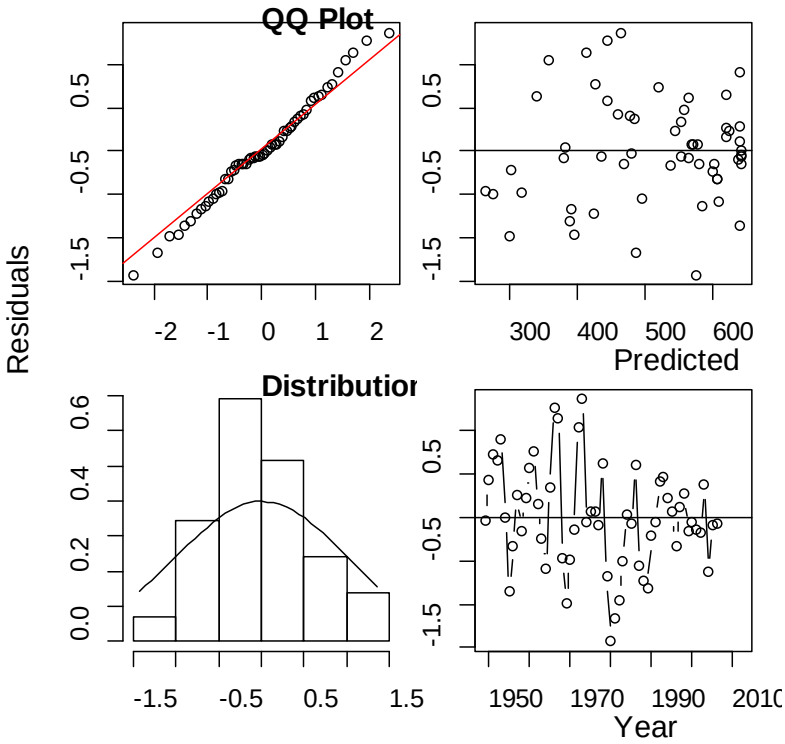

Figure 8. Residual diagnostic plots.

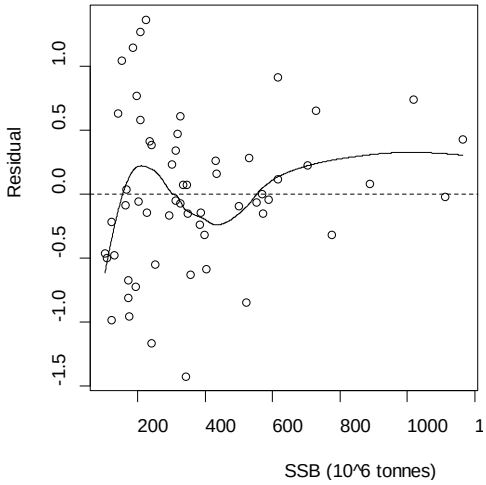

Figure 9. Residuals vs. stock size. The solid line indicates a loess (R, 2011) smooth of the residuals.

Recruitment predictions are shown in Figure 10. Predictions of the size of the 2004–2005 year classes were also quite different from the VPA estimates, similar to the BH model.

Comparison of retrospective predictions and current VPA estimates of recruitment are shown in Figures 11 and 12. Root mean squared errors (Table 1) are almost as large as the BH results.

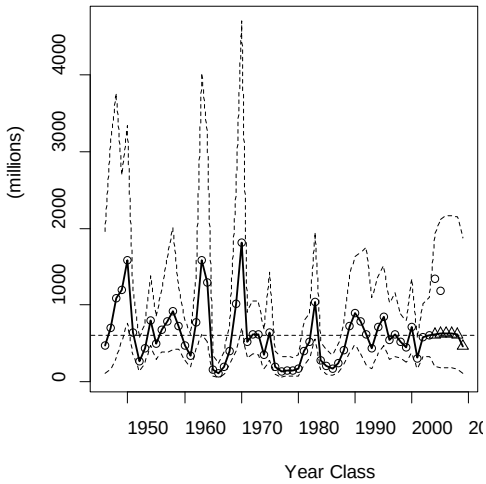

Figure 10. Model estimates (solid lines) and predictions (triangles) of recruitment for Barents Sea cod. The dashed lines indicate 95% prediction intervals. Circles are the converged VPA estimates, which are considered to be the best available estimates and used as a baseline.

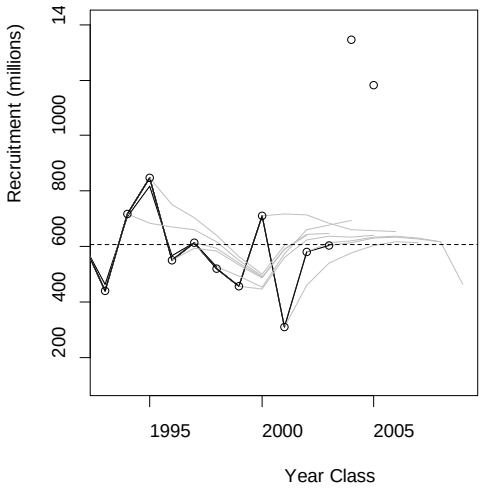

Figure 11. Retrospective estimates (solid lines) and predictions (grey lines), including three year forecasts, of recruitment. The model usually interpolated recruitment estimates, and the retrospective solid lines almost all overlap. Circles are the converged VPA estimates, which are considered to be the best available estimates and used as a baseline. The dashed line represents the average of the VPA estimates for 1946–2008; the time frame considered to be converged in the current assessment.

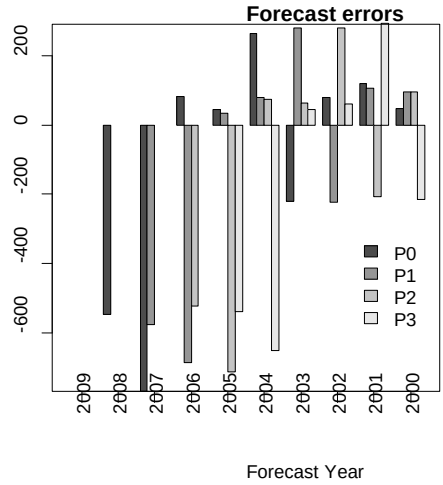

Figure 12. Differences between 2011 VPA estimates of recruitment to 2008 and retrospective stock–recruit model predictions for the current year (P0) and three year forecasts (P1–P3). Some P0–P4 differences could not be computed for 2006–2008 because some of the forecasts in these years were for recruitments later than 2008.

Discussion

It is usually difficult to separate observation and process errors. In this application all the variability was estimated to be process errors. It is not clear if this is appropriate. However, Minto (2011) implemented essentially the same model for this stock and found very similar results, so it does not seem that there are implementation errors. At this stage the method and results should be considered as "illustrative". Future research into the efficacy of the approach is required.

The method is not intended to be a competitor of recruitment models that incorporate specific ecological knowledge of the stock and additional survey and environmental data. However, it is an approach that requires only standard data and stock knowledge and can be easily applied to other stocks. It is also interesting to compare

the prediction accuracy of more complicated models with the simple approach proposed here.

P2 and P3 forecast accuracies did not include assessment variability of SSB. P2 and P3 forecasts are derived from SSB in the assessment year and the year previous to this, and these SSB's will not be converged and therefore have assessment variability that could affect the P2 and P3 forecast accuracies. However, the three most recent recruitment estimates were not used in the P1–P3 recruitment forecasts, and this likely overinflates the forecast standard deviations. A better approach to evaluate stock–recruit model forecast accuracies is to use retrospective stock–recruit estimates when deriving recruitment forecasts.

7 References

ADMB Project. 2009. AD Model Builder: automatic differentiation model builder. Developed by David Fournier and freely available from admb-project.org.

Alvarez-Fernandez, S., Lindeboom, H. and Meesters, E. 2012. Temporal changes in plankton of the North Sea: community shifts and environmental drivers. Marine Ecology Progress Series **462**, 21–38.

Armstrong, S. 1989. Combining forecasts: The end of the beginning or the beginning of the end? International Journal of Forecasting, **5**, 585–588.

Armstrong, S. 2001. Principles of forecasting. A handbook for researchers and practitioners. Dordrecht, The Netherlands: KAP.

Bates, J. M., Granger, C. W. J. 1969. The combination of forecasts. Operational Research quarterly, **20**, 451–468.

Bousquet, O. and Elisseeff, A. 2002. Stability and generalization. Journal of Machine Learning Research, **2**, 499–526.

Breiman, L. 1996. Bagging predictors. Machine Learning, **24**, 123–140.

Burnham, K. P., and Anderson, D.R. 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 2nd ed. Springer-Verlag. ISBN 0-387-95364-7.

Caley, M.J., Carr, M.H., Hixon, M.A., Hughes, T.P., Jones, G.P., and Menge, B.A. 1996. Recruitment and the local dynaics of open marine populations. Annu. Rev. Ecol. Syst. **27**: 477–500.

Clemen, R. T. 1989. Combining forecasts: A review and annotated bibliography. International Journal of Forecasting, **5**, 559–583.

de Gooijer, J. G., Abraham, B., Gould, A. and Robinson, L. 1985. Methods for determining the order of an autoregressive-moving average process: A survey. International Statistical Review A, **53**, 301–329.

Dingsør G.E., Bogstad B., Stiansen J.E., and Subbey S. 2010 How can we assess recruitment models for (age-3) NEA cod? //ICES AFWG 2010/WD:19.

Dutil, J.-D., and Brander, K. 2003. Comparing productivity of North Atlantic cod (*Gadus morhua*) stocks and limits to growth production. Fish. Oceanogr. **12**(4/5): 502–512.

Enberg, K., Jørgensen, C., Dunlop, E.S., Varpe, Ø., Boukal, D.S., Baulier, L., Eliassen, S., and Heino, M. 2011. Fishing-induced evolution of growth: concepts, mechanisms and the empirical evidence. Marine Ecology: no-no.

Enberg, K., rgensen, C., and Mangel, M. 2010. Fishing-induced evolution and changing reproductive ecology of fish: the evolution of steepness. Canadian Journal of Fisheries and Aquatic Sciences **67**(10): 1708–1719.

Evgeniou, T., Pontil, M. and Elisseeff, A. 2004. Leave one out error, stability, and generalization of voting combinations of classifiers. Machine Learning, **55**(1), 71–97.

Fox, C.J., Planque, B.P., and Darby, C.D. 2000. Synchrony in the recruitment time-series of plaice (*Pleuronectes platessa* L) around the United Kingdom and the influence of sea temperature. Journal of Sea Research **44**(1): 159–168.

Hibon, M. and Evgeniou, T. 2005. To combine or not to combine: selecting among forecasts and their combinations, International Journal of Forecasting, **21**, 15–24.

Huse, G., Fernö, A., and Holst, J.C. 2010. Establishment of new wintering areas in herring co-occurs with peaks in the 'first time/repeat spawner' ratio. Marine Ecology Progress Series **409**: 189–198.

Hutchings, J.A. 2005. Life-history consequences of overexploitation to population recovery in Northwest Atlantic cod (*Gadus morhua*). Can J Fish Aquat Sci **62**: 824–832.

Hutchings, J.A., and Rangeley, R.W. 2011. Correlates of recovery for Canadian Atlantic cod (*Gadus morhua*). Can J Zool **89**: 386–400.

ICES. 2008. Report of the Arctic Fisheries Working Group (AFWG). ICES CM 2008/ACOM:01.

ICES. 2011. Report of the Study Group on Recruitment Forecasting (SGRF), 17–20 October 2011, Copenhagen, Denmark. *In* ICES CM 2011/ACOM. p. 40.

ICES. 2012. Report of the Arctic Fisheries Working Group (AFWG). ICES CM 2012/ACOM:05.

Jobling, M. 1994. Fish bioenergetics. Chapman & Hal, London.

Kjesbu, O.S., Solemdal, P., Bratland, P., and Fonn, M. 1996. Variation in annual egg production in individual captive Atlantic cod (*Gadus morhua*). Can J Fish Aquat Sci **53**(3): 610–620.

Lorenzen, K. 2008. Fish Population Regulation Beyond Stock and Recruitment: The Role of Density-Dependent Growth in the Recruited Stock. Bulletin of Marine Science **83**(1): 181–196.

Lorenzen, K., and Enberg, K. 2002. Density-dependent growth as a key mechanism in the regulation of fish populations: evidence from among-population comparisons. Proc. R. Soc. Lond. B **269**: 49–54.

Makridakis, S., and Winkler, R. 1983. Averages of forecasts: Some empirical results. Management Science, **29**, 987–996.

Marteinsdottir, G., and Begg, G.A. 2002. Essential relationships incorporating the influence of age, size and condition on variables required for estimation of reproductive potential in Atlantic cod *Gadus morhua*. Mar Ecol Prog Ser **235**: 235–256.

McCullagh, P. and J.A. Nelder. 1989. Generalized Linear Models. Chapman and Hall: London. (mathematical statistics of generalized linear model).

Minto, C. 2011. Ecological inference from variable recruitment data. PH.D. thesis. Dalhousie University: Canada.

Munch, S.B., and Kottas, A. 2009. A Bayesian modelling approach for determining productivity regimes and their characteristics. Ecol Appl **19**(2): 527–537.

Myers, R.A. 1998. When do environment-recruitment correlations work? Rev Fish Biol Fisher **8**: 285–305.

Nash, R.D.M., and Dickey-Collas, M. 2005. The influence of life-history dynamics and environment on the determination of year-class strength in North Sea herring (*Clupea harengus* L.). Fish Oceanogr **14**(4): 279–291.

Needle, C.L.; O'Brien, C.M.; Darby, C.D.; Smith, M.T., 2003. Incorporating time-series structure in medium-term recruitment projections, in: Ulltang, Ø. *et al*. (2003). Fish stock assessments and predictions: integrating relevant knowledge: SAP Symposium held in Bergen, Norway 4–6 December 2000. Scientia Marina (Barcelona), **67**(Suppl. 1): pp. 201–209.

Neill, W.H., Miller, J.M., Van Der Veer, H.W., and Winemiller, K.O. 1994. Ecophysiology of marine fish recruitment: A conceptual framework for understanding interannual variability. Netherlands Journal of Sea Research **32**(2): 135–152.

O'Brien, C.M., Fox, C.J., Planque, B., and Casey, J. 2000. Fisheries: Climate variability and North Sea cod. Nature **404**(6774): 142–142.

Olsen, E.M., Ottersen, G., Llope, M., Chan, K.-S., Beaugrand, G., and Stenseth, N.C. 2011. Spawning stock and recruitment in North Sea cod shaped by food and climate. Proc R Soc B **278**(1705): 504–510.

Paulik, G.J. 1973. Studies of the possible form of the stock and recruitment curve. Rapp P-V Reun - Cons Int Explor Mer **164**: 302–315.

Payne, M.R., Hatfield, E.M.C., Dickey-Collas, M., Falkenhaug, T., Gallego, A., Gröger, J., Licandro, P., Llope, M., Munk, P., Röckmann, C., Schmidt, J.O., and Nash, R.D.M. 2009. Recruitment in a changing environment: the 2000s North Sea herring recruitment failure. ICES J Mar Sci **66**: 272–277.

Petitgas, P., Huret, M., and Léger , F. 2011. Identifying and monitoring limiting factors of recruitment: anchovy in the Bay of Biscay.

Planque, B., and Frédou, T. 1999. Temperature and the recruitment of Atlantic cod (*Gadhus morhua*). Can J Fish Aquat Sci **56**: 2069–2077.

R Development Core Team. 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Rose, K.A. 2005. Lack of relationship between simulated fish population responses and their life-history traits: inadequate models, incorrect analysis, or site-specific factors? Canadian Journal of Fisheries and Aquatic Sciences **62**(4): 886–902.

Rothschild, B.J. 1986. Dynamics of marine fish populations. Harvard University Press, Cambridge, USA.

Shepherd, J.G. 1997, Prediction of year-class strength by calibration regression analysis of multiple recruit index series. ICES Journal of Marine Science, **54**: 741–752. 1997.

Skjaeraasen, J.E., Nash, R.D., Korsbrekke, K., Fonn, M., Nilsen, T., Kennedy, J., Nedreaas, K.H., Thorsen, A., Witthames, P.R., Geffen, A.J., Hoie, H., and Kjesbu, O.S. 2012. Frequent skipped spawning in the world's largest cod population. Proceedings of the National Academy of Sciences of the United States of America **109**(23): 8995–8999.

Skjoldal, H.R. 2004. Fish stocks and fisheries in relation to climate variability and exploitation. *In* Natural resource system challenge: oceans and aquatic ecosystems. In: Encyclopedia of Life Supporting Systems (EOLSS), Developed under the auspices of the UNESCO. Edited *by* E. Wolanski. Eolss Publishers, Oxford, UK.

Solari, A.P., Martín-González, J.M., and Bas, C. 1997. Stock and recruitment in Baltic cod (*Gadus morhua*): a new, non-linear approach. ICES Journal of Marine Science: Journal du Conseil **54**(3): 427–443.

Stiansen, J.E., A. Aglen, B. Bogstad, H. Loeng, S. Mehl, O. Nakken, G. Ottersen and E. Svendsen. 2005. Including climate into the assessment of future fish recruitment, using multiple regression models. ICES CM 2005/O:25.

Subbey, S., JE. Stiansen, B. Bogstad, T. Bulgakova and O. Titov. 2008. Evaluating Recruitment Models for (Age 3) NEA Cod. //ICES AFWG 2008/WD:27.

Titov, O. V. 2001. System approach to the analysis of natural long-term changes of the Barents Sea ecosystem. Murmansk. PINRO. P. 119 (in Russian).

Titov O. V. 1999. Long-term cyclic succession of the Barents Sea ecosystem and prognostication of recruitment of commercial fish population // ICES CM 1999/O:01.- 17 pp.

Titov, O., Pedchenko, A. and Karsakov, A. 2005. 'Assessment of northeast Arctic cod and capelin recruitment from data on ecological situation in the Barents Sea I 2004-2005'. Working document #16 in: Report of the Arctic Fisheries Working Group', Murmansk, Russia, April 19–28, 2005. ICES C.M. 2005/ACFM:20, 564 pp.

Titov, O. 2011. Assessment of population recruitment abundance of northeast Arctic cod considering the environment data //ICES AFWG 2011/WD:23.

Titov, O. 2008 Assessment of population recruitment abundance of northeast Arctic cod considering the environment data //ICES AFWG 2008/WD.

Trevor Hastie, Robert Tibshirani and Jerome Friedman, "Elements of Statistical Learning: Data Mining, Inference and Prediction" Springer-Verlag, New York.

Trippel, E.A. 1995. Age at maturity as a stress indicator in fisheries. Bioscience **45**: 759–771.

Van Der Veer, H.W., Berghahn, R., and Rijnsdorp, A.D. 1994. Impact of juvenile growth on recruitment in flatfish. Netherlands Journal of Sea Research **32**(2): 153–173.

Vapnik, V. 1998. Statistical learning theory. New York, USA: John Wiley & Sons.

Vasilakopoulos, P., O'Neill, F.G., and Marshall, C.T. 2011. Misspent youth: does catching immature fish affect fisheries sustainability? ICES Journal of Marine Science **68**(7): 1525–1534.

Yang, Y. 2003. Regression with multiple candidate models: Selecting or mixing? Statistica Sinica, **13** (3), 783–810.

Yang, Y. 2004. Combining forecasting procedures: Some theoretical results. Econometric Theory, **20** (1), 176–222.

Zou, H. and Yang, Y. 2004. Combining time-series models for forecasting, International Journal of Forecasting, **20**, 69–84.

Annex 1: List of participants

Name	Address	Phone/Fax	Email
Noel Cadigan	Memorial University of Newfoundland Centre for Fisheries Ecosystems Research P.O. Box 4920 St John's NF A1C 5R3 Canada	Phone +1 709 772 5028 Fax +1 709 778 0603	Noel.Cadigan@mi.mun.ca
Jennifer Devine	Institute of Marine Research P.O. Box 1870 Nordnes 5817 Bergen Norway	Phone +47 55238588 Fax +47	jennifer.devine@imr.no
Giulia Gorelli	Institut de Ciències del Mar - CSIC Pg. Marítim de la Barceloneta, 37–49 E-08003 Barcelona Catalunya Spain	Phone +34 Fax +34	gorelli@icm.csic.es
Olav Kjesbu	Institute of Marine Research P.O. Box 1870 Nordnes 5817 Bergen Norway	Phone +47 55 238487 Fax +47 55 238584	olav.kjesbu@imr.no
Yuri Kovalev	Knipovich Polar Research Institute of Marine Fisheries and Oceanography(PINRO) 6 Knipovitch Street 183763 Murmansk Russian Federation	Phone +7 8152 472 469 Fax +7 8152 473 331	kovalev@pinro.ru
Richard Nash	Institute of Marine Research P.O. Box 1870 Nordnes 5817 Bergen Norway	Phone +47 55 23 68 55 Fax +47 55 23 85 31	Richard.Nash@imr.no
Isabel Palomera	Institut de Ciències del Mar - CSIC Pg. Marítim de la Barceloneta, 37–49 E-08003 Barcelona Catalunya Spain	Phone +34 Fax +34	isabel@icm.csic.es
Daniela Silveira Simao Oberserver	Institut de Ciències del Mar-CSIC Pg. Marítim de la Barceloneta, 37–49 E-08003 Barcelona Catalunya Spain		simao@icm.csic.es

Name	Address	Phone/Fax	Email
Andrey Sorokin	AtlantNIRO 5 Dmitry Donskogo Street RU-236000 Kaliningrad Russian Federation	Phone +7 Fax +7	Antares2002@rambler.ru
Jan Erik Stiansen	Institute of Marine Research P.O. Box 1870 Nordnes 5817 Bergen Norway	Phone +47 55 238 626 Fax +47 55 238 687	jan.erik.stiansen@imr.no
Samuel Subbey Chair	Institute of Marine Research P.O. Box 1870 Nordnes 5817 Bergen Norway	Phone +47 5523 5383 Fax +47 5523 8687	samuel.subbey@imr.no
Oleg Titov	Knipovich Polar Research Institute of Marine Fisheries and Oceanography(PINRO) 6 Knipovitch Street 183038 Murmansk Russian Federation	Phone +7 Fax +7	titov@pinro.ru