Department of Fisheries and Oceans
Canadian Stock Assessment Secretariat
Research Document 97/57

Ministère des pêches et océans
Secrétariat canadien pour l'évaluation des stocks
Document de recherche 97/57

EXAMINING THE RISK OF A STOCK STATUS CLASSIFICATION SYSTEM FOR EAST AND SOUTHEAST NEWFOUNDLAND HERRING STOCKS

William G. Warren

Science Branch
Department of Fisheries and Oceans
P.O. Box 5667
St. John's Newfoundland A1C 5X1

[1] This series documents the scientific basis for the evaluation of fisheries resources in Canada. As such, it addresses the issues of the day in the time frames required and the documents it contains are not intended as definitive statements on the subjects addressed but rather as progress reports on ongoing investigations.

Research documents are produced in the official language in which they are provided to the Secretariat.

[1] La présente série documente les bases scientifiques des évaluations des ressources halieutiques du Canada. Elle traite des problèmes courants selon les échéanciers dictés. Les documents qu'elle contient ne doivent pas être considérés comme des énoncés définitifs sur les sujets traités, mais plutôt comme des rapports d'étape sur les études en cours.

Les documents de recherche sont publiés dans la langue officielle utilisée dans le manuscrit envoyé au secrétariat.

Abstract

Wheeler and Winters (1996) presented a stock-specific model that relates recruitment of east and southeast Newfoundland herring to mature biomass and the overwintering/prespawning temperature and salinity. They also presented a stock status classification system which, perforce, uses solely the mature biomass, which might be reasonably projected. A potential problem with this system is the considerable variation about the recruitment-biomass relationship, of which much can be attributed to temperature and salinity. The risk, i.e. the probability of having the system select an inappropriate exploitation rate (f), is evaluated through the recruitment model and estimates of the distributions of salinity and temperature. Since temperature exhibits significant serial correlation, the temperature distribution is that of the temperature projected one year in advance from the current temperature. The method has been validated by its application to the existing time series of the four Newfoundland stocks and behaves within the bounds of expectation. Finally, it is used to project future recruitment for the four stocks at exploitation rates suggested by the classification system.

Résumé

Wheeler et Winters (1996) ont présenté un modèle de stocks établissant une relation entre, d'une part, le recrutement du hareng de l'est et du sud-est de Terre-Neuve et, d'autre part, la biomasse d'adultes et la température et la salinité pendant l'hiver et avant la fraie. Ils ont aussi établi un système de classification de la situation des stocks qui, forcément, repose uniquement sur la biomasse d'adultes, ce que l'on pourrait raisonnablement prévoir. Mais la variation importante de la relation recrutement-biomasse attribuable en grande partie à la température et à la salinité pourrait être un problème. Le risque, c'est-à-dire la probabilité que le système choisisse un taux inapproprié d'exploitation (f), est évalué par le biais du modèle de recrutement et d'estimations de la distribution de la salinité et de la température. Étant donné que la température montre une autocorrélation significative, sa distribution correspond aux températures prévues un d'avance par rapport aux températures actuelles. La méthode a été validée en l'appliquant à la série existante de données temporelles sur les quatre stocks de Terre-Neuve et a donné des résultats se situant dans les limites prévues. En dernier lieu, elle a été utilisée pour projeter le recrutement futur aux quatre stocks à des taux d'exploitation obtenus du système de classification.

Introduction

Based on Winters and Wheeler (1987), but with 10 years of additional data and egg production levels replaced by stock specific mature (5+) biomass, Wheeler and Winters (1996) presented environmentally dependent stock specific recruitment models of the form

$$R = MBe^{a+bBM+cT+dS}$$

where R denotes the recruitment, MB the mature biomass generating that recruitment, T and S the overwintering/prespawning temperature and salinity, respectively (Station 27 0-20m surface layer).

Wheeler and Winters (1996) used these stock-recruitment relationships to construct a stock-status classification sysyem. For each stock, the observed range of mature biomass was divided into four zones with the spawning-stock threshold level (i.e. level of maximum recruitment) as the key reference point; the corresponding biomass provided the boundary between zones 2 and 3. Based on work by Mace (1994) and Myers et al. (1994), the boundary between zones 1 and 2 was determined by the point on the stock-recruitment curve at which recruitment is half the maximum. The boundary between zones 3 and 4 was then arbitrarily set so that zones 2 and 3 had the same width of biomass. Exploitation rates (f) from 0 - 0.05 to $\geq$ 0.20 were suggested for zone 1 to zone 4, respectively.

One problem with basing a classification on solely the mature biomass is the considerable variation about the recruitment biomass relationship, of which much can be attributed to, in particular, temperature and, to some extent, salinity. If all three components, namely mature biomass, temperature and salinity, were known, a quite good estimate of recruitment seemingly could be made. While the mature biomass might reasonably be projected, there would remain the problem of forecasting the temperature and salinity. In what follows, the risk of using the stock classification system is explored when only past temperture and salinity data are available.

Recruitment Models

East and southeast Newfoundland herring stocks comprise four stock complexes, White Bay - Notre Dame Bay, Bonavista Bay -Trinity Bay, St. Mary's Bay - Placentia Bay and Fortune Bay. For each stock the parameters of the above relationship, estimated by least squares on the linearized form,

$$\log(R/MB) = a + bMB + cT + dS$$

are

Location	a	b	c	d
White Bay-Notre Dame Bay	-158.2340	-0.013226	0.3604525	5.1351027
Bonavista Bay-Trinity Bay	-117.0064	-0.023044	0.3367788	3.8460613
St. Mary's Bay-Placentia Bay	-124.5046	-0.065150	0.2649980	4.0930709
Fortune Bay	- 78.0324	-0.109228	0.3225298	2.2691675

The residual mean squares are 0.8940618, 1.5457306, 1.5572237 and 3.0524283, respectively, (standard deviations of 0.9455, 1.2433, 1.2479 and 1.7471).

There are several reasons for these parameter estimates to differ somewhat from those given by Wheeler and Winters (1996); additional years of data have been used, with revised estimates of recruitment and mature biomass, and age 2 rather than age 3 recruitment. Also, actual recruitments (in thousands) and mature biomass (in thousands of tonnes) have been employed, whereas the rounded ratios used previously sometimes resulted in zero, so that 1 was then added to the ratios prior to taking logarithms.

For the most part, all regression parameters are significant (5% level or better). Sequential F-tests are as follows:

Location	Test	F_{df}	p
White Bay-Notre Dame Bay	MB	$F_{1,27}$=11.2	0.002
	T|MB	$F_{1,26}$ = 11.8	0.002
	S|MB,T	$F_{1,25}$ = 17.7	0.0003
Bonavista Bay-Trinity Bay	MB	$F_{1,27}$=20.5	0.0001
	T|MB	$F_{1,26}$=9.27	0.005
	S|MB,T	$F_{1,25}$=5.77	0.024
St. Mary's Bay-Placentia Bay	MB	$F_{1,28}$=4.20	0.050
	T|MB	$F1,27$=5.09	0.032
	S|MB,T	$F_{1,26}$=7.89	0.009
Fortune Bay	MB	$F_{1,26}$=2.23	0.15
	T|MB	$F_{1,25}$=4.34	0.048
	S|MB,T	$F_{1,24}$=1.26	0.28

The relationships appear to be much stronger in White Bay - Notre Dame Bay and Bonavista Bay - Trinity Bay than in St. Mary's Bay - Plancentia Bay and, in particular, Fortune Bay. In Fortune Bay the relationship is only marginally significant, and salinity appears to have no predictive power. Note that TMB (i.e. T after fitting MB) is significant while T on its own is not. The marginal significance of MB and T arises when they are used jointly, implying that the effect of MB is masked by T and vice versa. Although the omission of S from the Fortune Bay regression would appear to be justified, for consistency, it has been included in what follows.

One might conjecture that the strength of the relationships relates to the location of the areas relative to Station 27. White Bay - Notre Dame Bay and Bonavista Bay - Trinity Bay are on the east coast of Newfoundland, north of Station 27, but with conditions at all three sites controlled very much by the Labrador current. St Mary's Bay - Placentia Bay is on the southern coast as is Fortune Bay, but with the latter further to the west. Both sites should be less affected by the Labrador current; in particular, it seems feasible that the salinity in Fortune Bay would be influenced by the outflow of the St Lawrence. The weakening of the relationships may be due, therefore, not so much to a lack of dependence on temperature and salinity but the inappropriateness of Station 27 to represent conditions at these sites.

Forecasting

The time series of salinities (29 years) shows negligible serial correlation. Accordingly, the only predictor of salinity would be a value chosen at random from the estimated probability distribution of salinities.

In contrast, the time series of temperatures exhibits strong serial correlation. Rather than applying formal time series methods, it would seem sufficient to regress the temperature in year $i+1$ on the temperature in year i. The difficulty here lies not so much in the regression but in the pattern of residuals. Basically, if the temperature in year i is low then it is far more likely that temperature in year $i+1$ will increase over that of year i than drop below it. Likewise, if the temperature in year i is relatively high, it is more likely that the temperature in year $i+1$ will decrease below that of year i than fall above it. Thus, rather than being symmetric, the distributions of temperature in year $i+1$, given the temperature in year i, would appear to pass progressively from being distinctly positively skew, for low temperature in year i, to being distinctly negatively skew, for high temperature in year i.

It is also reasonable to assume that there are finite bounds to the overwintering/prespawning temperatures that can arise. Let these be t_0 and t_m and let $w = t_m - t_0$. If we assume that t has a Pearson Type IV distribution, i.e. t has probability density

$$(t - t_0)^{\alpha-1}(t_0 + w - t)^{\beta-1}/w^{\alpha+\beta-1}B(\alpha, \beta)$$

where α, $\beta > 0$ and $B(.,.)$ denotes the beta function, then the transformed variate, $z = (t - t_0)/w$ has the standard beta distribution

$$z^{\alpha-1}(1 - z)^{\beta-1}/B(\alpha, \beta).$$

Let x denote the temperature in a year and t the temperature in the following year. We assume that $E(t) = a + bx$. Futher, we suppose that $Var(t)$ is constant, i.e. independent of x. This is a strong assumption but one that seems reasonable and is analogous to the usual assumption of constant variance in ordinary least-squares regression. If we make the transformation $z = (t - t_0)/w$ then $E(z) = (a - t_0 + bx)/w$, which, for the ith year $x = x_i$ we may set equal to $\alpha_i/(\alpha_i + \beta_i)$. Further, $Var(z) = (b/w)^2 Var(y) = \sigma_z^2$, say, which we may set equal to

$$\frac{\alpha_i\beta_i}{(\alpha_i + \beta_i)^2(\alpha_i + \beta_i + 1)}$$

We thus have 5 parameters, t_0, w, a, b and σ_z^2 to estimate, with maximum likelihood being the obvious choice of method.

The logarithm of the likelihood can be written

$$L = \sum_i [\alpha_i \log(z_i) + (\beta_i - 1)\log(1 - z_i) - \log(\Gamma(\alpha_i)) - \log(\Gamma(\beta_i)) + \log(\Gamma(\alpha_i + \beta_i))]$$

Given t_0 and w (and, therefore, the z_i) it is a relatively easy matter to obtain the m.l.e. of a, b and σ_z^2. One could then search for the overall m.l.e. by repeating the process with different values of t_0 and w. It turns out, however, that the unconstrained m.l.e. of the upper bound to temperature is unreasonably large. Very high temperatures would be predicted albeit with low probability and a good fit to the data should be obtained, but in using this estimate in a Monte Carlo simulation with a large number of realizations, a few such values would inevitably arise and lead to unrealistically high projected recruitments. Accordingly, final upper and lower temperaure bounds have been chosen somewhat arbitrarily. It is clear that the lower bound cannot be greater than the lowest value (0.01) observed in the series, nor can the upper bound be less than the highest observed value (11.40). Accordingly t_0 was set at -1.0 and w at 14.0 (i.e. t_m=13.0). These values then lead to conditional m.l.e. of a=2.245, $b = 0.445$ and $\sigma^2 = 0.028$. (Note that these a and b are very close to the ordinary least-squares estimates).

How viable are these estimates? The following study was undertaken on a slightly shorter time series prior to the data for the most recent years being made available; the conditional m.l.e. were $a = 2.4$ and $b = 0.49$ with σ^2 unchanged at 0.028. For the first 28 temperatures the α_i and β_i were calculated and the corresponding estimated probability densities of t were plotted. The observed value of z (i.e. x_{i+1} transformed) was superimposed on these plots. As would be hoped, about 1/3 of these were found to be close to the modal value of the distribution, about 1/3 were clearly to the left of this and about 1/3 clearly to the right. There was only one case where the value was in the extreme upper tail; this occured when $x_i = 6.00$ increased to 11.40 in the following year, by far the greatest jump in the series. The most extreme case where the observed value was in the lower tail was still within the bounds of expectation and corresponded to the somewhat surprising drop from 3.45 in one year to 0.01 in the next. Notwithstanding, the model does have some predictive ability.

The empirical distribution of salinites exhibits negative skewness. Again it is reasonable to assume that there are upper and lower bounds to the possible values of salinity. As with temperature, appropriately transformed salinities can be modelled by a standard beta distribution and, again as with temperature, unconstained maximum likelihood estimation can lead to unrealistically high salinities (and thence, in Monte Carlo simulations, to unrealistically high recruitments). A plot of the order statistics of salinity suggests lower and upper bounds of 31.5 and 32.3 (range = 0.8). With these values assumed, the m.l.e of the beta distribution parameters are $\alpha = 2.95$ and $\beta = 1.74$. A plot of the resulting probability density of salinity superimposed on a histogram of the salinity data showed reasonable agreement, given that bimodality was suggested by the lack of salinity values around 31.8. However, bimodality can often appear by chance in samples of this size of a strictly unimodal variate and, a priori, there would appear to be no reason to suspect that the distribution of salinity would be anything other than unimodal.

Given projected biomass and current temperature, we are now in a position to project the distributions of future temperature and salinity and, hence, recruitment. (The time series of temperatures and salinities suggest that these are independent). An IML program for doing this is appended. The same temperatures and salinities are used for all regions. Note that after generating $\log(R/MB)$ via a+bMB+cT+dS, a normal random variable (mean zero, variance = the residual mean square for the region; see above) is added.

Validation

To evaluate the viability of the above model, 1000 realizations of recruitment were generated for each of the four stocks and each of the first 28 years of the data (29 for SMP-PB, 27 for FB). The temperatures and salinities of all years are, of course, correct however the projected biomasses and recruitments of the final few years are regarded as less reliable. For each of the 112 cases the probabilities of recruitment falling below the observed values was then estimated. Now recruitment, R, and therefore $p = \int_0^R dF(R)$ are random variables. However, whatever the form of $F(R)$, in theory p has a uniform distribution. We may therefore compare the 112 probabilities with a uniform distribution. We obtain

p	0-0.1	0.1-0.2	0.2-0.3	0.3-0.4	0.4-0.5	0.5-0.6	0.6-0.7	0.7-0.8	0.8-0.9	0.9-1.0
fcy	6	12	20	11	8	12	12	9	11	11

A chi-squared test of the null hypothsis gives $\chi^2 = 10.86$ on 9 d.f. (prob ≈ 0.3) indicating that the hypothesis is quite acceptable. The overall mean of the probabilities is 0.4875, close to the expected value of 0.50. The means by sites are 0.5125, 0.4687, 0.4792 and 0.4898, suggesting that the performance within sites in close to expectation. There are, however, years where the average of the probabilities is significantly high or significantly low. The most extreme of these is 1968 with a mean of 0.9710. Here the temperature increased from 6.00 to 11.40, a greater jump than expected, and is also associated with one of the higher salinities (32.14). Under such circumstances the observed value would be in the upper tail of the probability distribution of the projected recruitment. The other years with significantly high average probabilities are associated with the higher salinities, although the temperatures appear to be in the range of expectation. The years of significantly low average probabilities are associated with either low salinities or with temperature drops when, on the basis of the time series, increases would seem more likely. Overall, the approach appears to give some improvement over what would be obtained by ignoring temperature.

Application

Monte Carlo projections (999 realizations) have been made of age 2 recruitment on the basis of the 1996 XSA mature biomass and temperature (7.23°), firstly with no fishing mortality and, secondly, with a couple of assumed levels of fishing mortality, chosen on the basis of the proposed stock classification system (Wheeler and Winters 1996). Plots of the empirical cumulative distributions evaluated at the 1, 5, 10, 15, 20, ... 80, 85, 90, 95 and 99 percentiles on log probability paper indicated that the distributions are essentially log normal. (This would follow automatically from the assumption that the residuals about the regression of log(R/MB) on MB, T and S being normally distributed, but is presumably only an approximation because of the beta distributions assumed for T and S).

To estimate the p^{th} percentile of the recruitment distribution one may use the $1000p$ order statistic of the Monte Carlo realizations or the fitted lognormal distribution. In the latter case, if m and s^2 are the estimated lognormal mean and variance, the p^{th} percentile is estimated as $\exp(m + probit(p)s)$. Conversely, to estimate the probability that the recruitment would fall below a specified value, one would calculate $\Phi(\log(R) - m)/s)$, where $\Phi(.)$ denotes the cumulative probability of the standard normal distribution, available in most computer software (as is the probit function).

The results are:

Estimates of age 2 recruitment given no fishing mortality ($\times$ 1000).

1. Based on 999 Monte Carlo realizations.

Location	Median	Interquartile Range
WB-NDB	86700	29900 - 246000
BB-TB	33500	11800 - 108000
SMB-PB	1800	660 - 5250
FB	5500	1400 - 19400

2. Based on fitted lognormal distribution.

Location	Median	Interquartile Range
WB-NDB	84100	30100 - 235000
BB-TB	35200	12100 - 102000
SMB-PB	1800	650 - 5000
FB	5300	1400 - 19900

The lognormal parameters are

Location	Mean(m)	S.D.(s)
WB-NDB	11.4339505	1.5244559
BB-TB	10.468138	1.5779094
SMB-PB	7.502271	1.5106592
FB	8.5717775	1.9660044

Estimates of age 2 recruitment given prescribed rates of fishing mortality.

1. Based on 999 Monte Carlo realizations.

Location	f	Median	Interquartile Range		
WB-NDB	0.05	80900	28700	-	245000
	0.1	84200	29700	-	250000
BB-TB	0.05	32700	11100	-	113000
	0.1	29300	10100	-	93000
SMB-PB	0.2	2500	860	-	7000
	0.4	2500	1200	-	7100
FB	0.2	8700	2300	-	34400
	0.4	12300	3200	-	44000

2. Based on fitted lognormal distributions.

Location	f	Median	Interquartile Range		
WB-NDB	0.05	81500	28800	-	230000
	0.1	84000	29900	-	296000
BB-TB	0.05	34000	11000	-	105000
	0.1	29700	10000	-	88500
SMB-PB	0.2	2400	870	-	6800
	0.4	2500	880	-	7000
FB	0.2	8500	2300	-	31900
	0.4	11900	3200	-	43500

The lognormal parameters are

Location	f	Mean(m)	S.D.(s)
WB-NDB	0.05	11.308180	1.5402394
	0.1	11.339051	1.5303388
BB-TB	0.05	10.434474	1.6675108
	0.1	10.299697	1.6180379
SMB-PB	0.2	7.799038	1.55289364
	0.4	7.819289	1.5386997
FB	0.2	9.052244	1.9522193
	0.4	9.381538	1.9244995

Note that with WB-NDB and BB-TB the effect of such low fishing mortalities is masked by the degree of uncertainty in the estimates. With SMB-PB, as expected, fishing appears to increase recruitment slightly, although the chances of low recruiment remain relatively high. With FB, again as expected, fishing results in a somewhat greater increase in recruitment; indeed, even with $f = 0.4$ the mature biomass remains in the good - very good region.

Discussion.

To sum up, the method appears to behave within the bounds of expectation. It results in a somewhat, but not spectacularly, tighter distribution of projected recruitment than would be obtained from the use of mature biomass alone. The main weakness lies in the inability to forecast salinity. Also, results will be relatively poor if the change in temperature is greater than, or in the opposite direction to, that which would be anticipated. There is reason to believe that recruitment is goverened by temparuture and salinity as well as mature biomass and prediction would be improved considerably if one had site specific data on temparature and salinity. While station 27 data might be adequate for White Bay - Notre Dame Bay and Bonavista Bay - Trinity Bay, it is less so for St. Mary's Bay - Placentia Bay and almost irrelevant for Fortune Bay.

Finally, the procedure can be readily updated as further information comes to hand, This includes not only site specific data and data from additional years but also any other input into appropriate upper and lower limits of temperature and/or salinity.

References

Mace, P.M. 1994. Relationship between common biological reference points used as thresholds and targets of fishereis management strategies. Can. J. Fish. Aquat. Sci., 51:110 122.

Myers, R.A., Rosenberg, A.A., Mace, P.M., Barrowman, N. and Restrepo, V.R. 1994. In search of thresholds for recruitment overfishing, ICES J. Mar. Sci., 51:191-205.

Wheeler, J.P. and Winters, G.H. 1996. Newfoundland east and southeast coast herring - an assessment of stocks to the spring of 1995. DFO Atl. Fish. Res. Doc. 96/63.

Winters, G.H. and Wheeler. J.P. 1987. Recruitment dynamics of spring-spawning herring in the Northwest Atlantic. Can. J. Fish. Aquat. Sci., 44:882-900.

APPENDIX

IML Computer program for the projection of the distribution of recruitment.

```
start sml; reset noprint;
rc=j(rzn,4,0); cnt=0; tv=j(rzn,1,0); sv=tv; r=j(4,1,0);
do j=1 to rzn;
p=uniform(sd);
z=(a-u+b*t0)/w;
alph=z*(z*(1-z)/v-1);
beta=alph*(1/z-1);
run invbeta;
t=w*xp+u; tv[j,1]=t;
p=uniform(sd);
alph=aa; beta=bb;
run invbeta;
s=ww*xp+uu; sv[j,1]=s;
do i=1 to 4;
r[i,1]=(1||bm[i,1]||t||s)*pms[,i];
r[i,1]=log(bm[i,1])+r[i,1]+se[i,1]*normal(sd);
end;
r=exp(r);
rc[j,]=r`;
end;
rx=rc;
do i=1 to 4;
rx[rank(rx[,i]),i]=rc[,i];
end;
finish;

start invbeta; reset noprint;
yp=probit(1-p);
h=2/(1/(2*alph-1)+1/(2*beta-1));
lam=(yp**2-3)/6;
wx=yp*sqrt(h+lam)/h-(1/(2*beta-1)-1/(2*alph-1))*(lam+5/6-2/(3*h));
xp=alph/(alph+beta*exp(2*wx));
finish;
```

REQUIRED INPUT

rzn - the number of Monte Carlo realizations (999)
a - the m.l.e. of the intercept of the temperature regression (2.4)
b - the m.l.e. of the slope of the temperature regression (0.49)
u - the assumed lower bound of temperature (-1)
w - the assumed range of temperature (14 = 13-(-1))
v - the m.l.e of the temperature variance (.028)
aa - the m.l.e. of alpha of the salinity distribution (3.9)
bb - the m.l.e. of beta of the salinity distribution (2.1)
uu - the assumed lower bound of salinity (31.5)
ww - the assumed range of salinity (0.8 = 32.3 - 31.5)
sd - any number to initialize the random number generator
bm - the projected biomasses of the 4 regions as a column vector
pms - the estimated parameters of the recruitment models
(with regions as columns i.e. the transpose of the array on
p.1 of the report)
se - the standard errors of the recruitment models as a column vector
t0 - the temperature used for prediction

The distribution of projected recruitments appears as a rzn x 4 array, rc, with regions as columns. The columns are sorted in ascending order in the array rx. Realized temperatures and salinities are stored in vectors tv and sv, should they be required.