ICES Journal of Marine Science, 58: 622–632. 2001
doi:10.1006/jmsc.2001.1068, available online at http://www.idealibrary.com on IDEAL®

Intercalibration of North Sea International Bottom Trawl Surveys by fitting year-class curves

A. J. R. Cotter

Cotter, A. J. R. 2001. Intercalibration of North Sea International Bottom Trawl Surveys by fitting year-class curves. – ICES Journal of Marine Science, 58: 622–632.

Intercalibration of trawl surveys is needed whenever trawling technique, the survey vessel, the season, or the localities of trawling are altered. Statistical modelling offers a generally cheaper method of estimating intercalibration factors than comparative trawling trials. Much variability of whole-survey population abundance indices transformed to natural logarithms can be explained using only the year-class strengths and a coefficient of total mortality, Z, for the species. This modelling approach was therefore used to intercalibrate surveys forming part of the International Bottom Trawl Survey of the North Sea between 1977 and 1997 for four commercially important galoid species, cod, haddock, whiting, and Norway pout. An age-related factor was included to allow for apparently lower catchabilities of young fish. The models fitted satisfactorily, permitting intercalibration factors to be estimated with standard errors. No indications of changes in Z were found over the period or over different year classes for any of the species. Residual errors were positively correlated among-ages-within-years for each survey. Residual degrees of freedom were therefore reduced using an information measure before testing factors in the model for significance or estimating standard errors. A method for comparing the relative precisions of the different surveys given the fitted model is also described.

Key words: trawl surveys, abundance indices, intercalibration, year-class curves, North Sea, cod (*Gadus morhua*), haddock (*Melanogrammus aeglefinus*), whiting (*Merlangius merlangus*), Norway pout (*Trisopterus esmarki*).

Received 27 July 2000; accepted 2 March 2001.

A. J. R. Cotter: Centre for Environment, Fisheries and Aquaculture Science, Lowestoft, England, NR33 0HT. e-mail: a.j.cotter@cefas.co.uk

Introduction

Four important factors which shuld be consistent over the time and space covered by a trawl survey are: (1) the trawl gear and towing technique, (2) the vessel, (3) the season(s) of towing, and (4) the stations or sampling strata where tows are located (Ona and Godø, 1990; Godø, 1994; Mitson, 1995; Munro, 1998). In practice, changes can seldom be avoided as the years pass, for example when trawling methods must be re-standardized to match those used by other surveys, when the vessel must be replaced, or when trawling stations are obstructed. Maintaining comparability of survey methods over time and across geographic regions poses particular problems for multi-vessel surveys such as the quarterly International Bottom Trawl Survey (IBTS) of the North Sea (ICES, 1996a, 1998a; Heessen *et al.*, 1997). This paper is about estimation of intercalibration factors needed to adjust survey abundance indices to allow for inconsistencies of method. The term "abundance index" is the conventional name for the mean catches per unit effort (cpues) estimated by surveys and is used here without questioning the assumptions inherent in inferring stock abundance from cpue (Paloheimo and Dickie, 1964; Myers and Stokes, 1989; Swain and Sinclair, 1994; Pennington and Godø, 1995).

Comparative trawling experiments provide a practical way of estimating intercalibration factors but they are expensive due to the high costs of operating vessels at sea. Reviews by Pelletier (1998) and Wilderbuer *et al.* (1998) of many reported experiments indicate that tens or, better, hundreds of tows are required for satisfactory precision. Efficiency is seriously impaired if fish stocks are low because small or zero catches contribute little to the precision (Pelletier, 1998). Estimated intercalibration factors may be biased for application outside the geographic locality of the trials (Pelletier, 1998), for example if the noisier vessel of a pair being compared disturbs fewer fish in deep water. A mismatch of season might also cause bias.

Statistical modelling of individual catches provides a relatively inexpensive, theoretical method for estimating

1054–3139/01/030622+11 $35.00/0

intercalibration factors (Sparholt, 1990; ICES, 1992; Cotter, 1993; Munro, 1998). One problem is that many individual, potentially interacting factors may serve to predict catch sizes, e.g. year, region, depth, time of day, etc. aside from the ship- and gear-related factors. The best, most durable model is therefore difficult to identify. Another problem is dependence among survey catch data (Myers and Cadigan, 1995). This can arise among catches at nearby stations due to associations of weather, locality, time of year, etc., and among fish of different categories (species, age, etc.) within single catches, due to the fact that a trawl collects clusters of fish rather than random samples (Pennington and Vølstad, 1994; Cotter, 1998). Over-fitted models, satisfactory in one year but not n the next, are a likely consequence when t or F tests are applied to assess the significance of factors in a model without first reducing residual degrees of freedom (d.f.) to allow for dependences.

Since stock assessments usually only consider whole-survey abundance indices, these indices can be intercalibrated by modelling in preference to individual catches without loss of any information likely to be needed for an assessment. The approach is simply to estimate the differences between survey indices without attempting to model the factors which might have caused the differences. The advantages are that identification of a suitable model is greatly simplified and computing is easier because the abundance indices are substantially fewer in number and less variable than the individual catches. A suitable model is that variously described by Jensen (1939), Ricker (1975) and King (1995) which will here be referred to as a year-class curve, i.e.

$$\ln N_{a,c} = \ln N_{0,c} - Za$$

This logarithmic model can explain much variation in abundances, N, of a year class (or cohort) c of a species with increasing age, a, simply by fitting an initial abundance plus a single coefficient of instantaneous total mortality, Z. One or two extra terms can easily be added to represent the intercalibration factors, as well as non-linearity if necessary. The minimal number of parameters in the model conserves residual d.f. for precise estimation. More complicated models have been described, e.g. that used by Cook (1997) to compare survey indices with stock assessments, but they are more extravagant with d.f.

The present paper illustrates application of year-class curves to intercalibrate nine IBTS survey indices for four species of gadoids in the North Sea, namely cod (*Gadus morhua*), haddock (*Melanogrammus aeglefinus*), whiting (*Merlangius merlangus*) and Norway pout (*Trisopterus esmarki*). Each survey was assumed to be fishing the same stocks and therefore to be estimating the same year-class strengths and Z or each species. This is consistent with the approach used for stock assessments (e.g. ICES, 1998b). Correlations of residuals between ages in each annual set of survey indices were allowed for in confidence limits and significance tests by reducing d.f. in accordance with an estimated measure of information. Useful by-products of the fitted models were the opportunity to compare relative survey precisions, and joint estimates of year-class strengths. The latter are being studied separately.

Materials and methods

The surveys

Operational details of each survey, the five-character abbreviated survey name used here, and the age-groups of fish whose abundance indices were modelled are given in Table 1. Most of the trawling and catch-sampling methods were standardized (ICES, 1996a, addendum) but four surveys (EGRT3, SABD3, AHER2 and AGOV2) did not use a GOV (grande overture verticale) trawl of the standard IBTS specification and the German survey vessel changed in 1985 (Ehrich, 1991). Geographic coverage for the purpose of estimating the indices varied as shown in Figure 1. Each national survey trawled once annually in each of 44–85 ICES rectangles (dimension: half a degree of latitude by one degree of longitude). The IGOV, however, was a collaborative effort by several nations and vessels over the whole North Sea, Skagerrak and Kattegat with more than 300 trawl stations distributed over approximately 160 rectangles. Collectively, these surveys formed part of the IBTS.

The data

Time-series of survey abundance indices were taken from ICES North Sea demersal stock assessment reports or, if the data were not tabulated there, from local CEFAS archives. Units given in the ICES reports could not all be reconciled confidently among the surveys, apparently due to re-scaling by unstated powers of ten. (The units are not needed for stock assessments.) For this reason, all original data were converted to "fish h^{-1}" as deduced from the stated units, then multiplied by whichever power of ten would lead to a set of intercalibration factors in the order of unity. The powers of ten used for each species and survey are shown in Table 1.

A few annual indices were missing. Those for 0-group fish were excluded because this age-group was poorly sampled by the IBTS. Those for the oldest age-groups were excluded if many of the indices were zero due to scarcity; "plus-group" fish were excluded in case the abundances of aggregated age-groups had different statistical properties from those of single year classes. The

Table 1. Intercalibration of North Sea IBTS: survey names and abbreviations used here, operational details, and the species and age-groups whose abundance indices were modelled with, beneath in parentheses, the source of data and the scaling factor applied to the original values after converting to numbers h^{-1} (see text). GOV=standard IBTS trawl; *=other trawls. Sources: 1=ICES (1996b); 2=ICES (1998b); 3=CEFAS.

Survey and abbreviation	Qtr	Vessel	Years	Trawl type	Cod	Haddock	Whiting	Norway pout
English EGRT3	3	Cirolana	1977–1991	Granton*	1–7 (2,1 000)	1–7 (2,1 000)	1–7 (2,1 000)	1–4 (3,100)
English EGOV3	3	Cirolana	1992–1997	GOV	1–7 (3^1,1)	1–7 (3^1,1)	1–7 (3^1,1)	1–4 (3^1,100)
Scottish SABD3	3	Scotia	1982–1996	Aberdeen*	1–7 (2,1 000)	1–7 (2,1 000)	1–7 (2,1 000)	1–4 (2,10)
Inter-national IGOV1	1	Various	1983–1997	GOV	1–6 (2,1)	1–6 (2^2,1 000)	1–6 (2^2,1 000)	1–3 (2,1)
German AHER2	2	Anton Dohrn	1983–1984	180′ Herring*	1–6 (1,1 000)	1–5 (3,1 000)	1–6 (3,1 000)	—
German AGOV2	2	Anton Dohrn	1985–1986	GOV with bobbin groundrope*	1–6 (1,1 000)	1–5 (3,1 000)	1–6 (3,1 000)	—
German WGOV2	2	Walther Herwig II	1987–1995	GOV	1–6 (1,1 000)	(1–5) (3,1 000)	1–6 (3,1 000)	—
Scottish SGOV2	2	Scotia	1991–1996	GOV	1–6 (2,1)	1–6 (2,1 000)	1–6 (2,1 000)	—
English EGOV4	4	Cirolana	1991–1996	GOV	1–6 (2,1)	1–6 (2,1 000)	1–6 (2,1 000)	1–4 (3,1)

[1]Source (3) was used in preference to (2) because indices in (2) were already corrected for the change of trawl from EGRT3 to EGOV3.
[2]The ICES Working Group tabulated these IGOV1 indices against the year prior to collection (ICES, 1996b, para 3.3.3).

few zero abundance indices remaining in the data set were also omitted to permit log transformation of the data without the complication and possibly worse bias of an added constant.

The model

Year-class curves with added terms for intercalibration and non-linearity factors were fitted to each species separately. The model was

$$\ln N_{a,c,s}=\ln N_{0,c}+Za+S_s+\beta\ \ln(a+1)+\varepsilon_{a,c,s} \qquad (1)$$

$N_{a,c,s}$ is the observed abundance index for fish aged a from year class c obtained by survey s, a being measured arbitrarily from 1 January of year c in years and months (the months being reckoned as decimals of a year). $N_{0,c}$ is the initial abundance index for the year class, and S_s is the intercalibration factor (as a natural logarithm) for survey s relative to one of the others, taken arbitrarily as EGOV3. S_s includes gear-, ship-, season-, and region-related factors, depending on which survey is being compared to EGOV3. $\varepsilon_{a,c,s}$ is the residual random deviation from the fitted model where $E(\varepsilon_{a,c,s})=0$, $Var(\varepsilon_{a,c,s})=\sigma^2$. The ln(a+1) term was added to the model to allow for apparently lower vulnerabilities or availabilities of young fish to the survey trawls (Ricker, 1975; Godø and Sunnanå, 1992; King, 1995) despite removal of 0-group indices from the analysis. It was found significant for all species when tested with the method described below. Note that age and survey are being treated as fixed rather than random factors. Equation (1) was fitted by ordinary least squares using indicator (0 or 1) variables for each year class, and for each survey except EGOV3, which was denoted by zeroes against all the other surveys. This provided immediate estimates of all year-class abundances, intercalibration factors, and the common Z.

Analysis of relative residual variances

Having fitted Equation (1), the relative precisions of the surveys were assessed, first by separating the residuals of each survey in turn from those of all the others, and second by calculating two estimates of σ^2, one, denoted $\hat{\sigma}_s^2$, from the single survey and one, denoted $\hat{\sigma}_r^2$, from the remaining surveys. The mean residual is zero in both cases as a result of fitting intercalibration factors by least

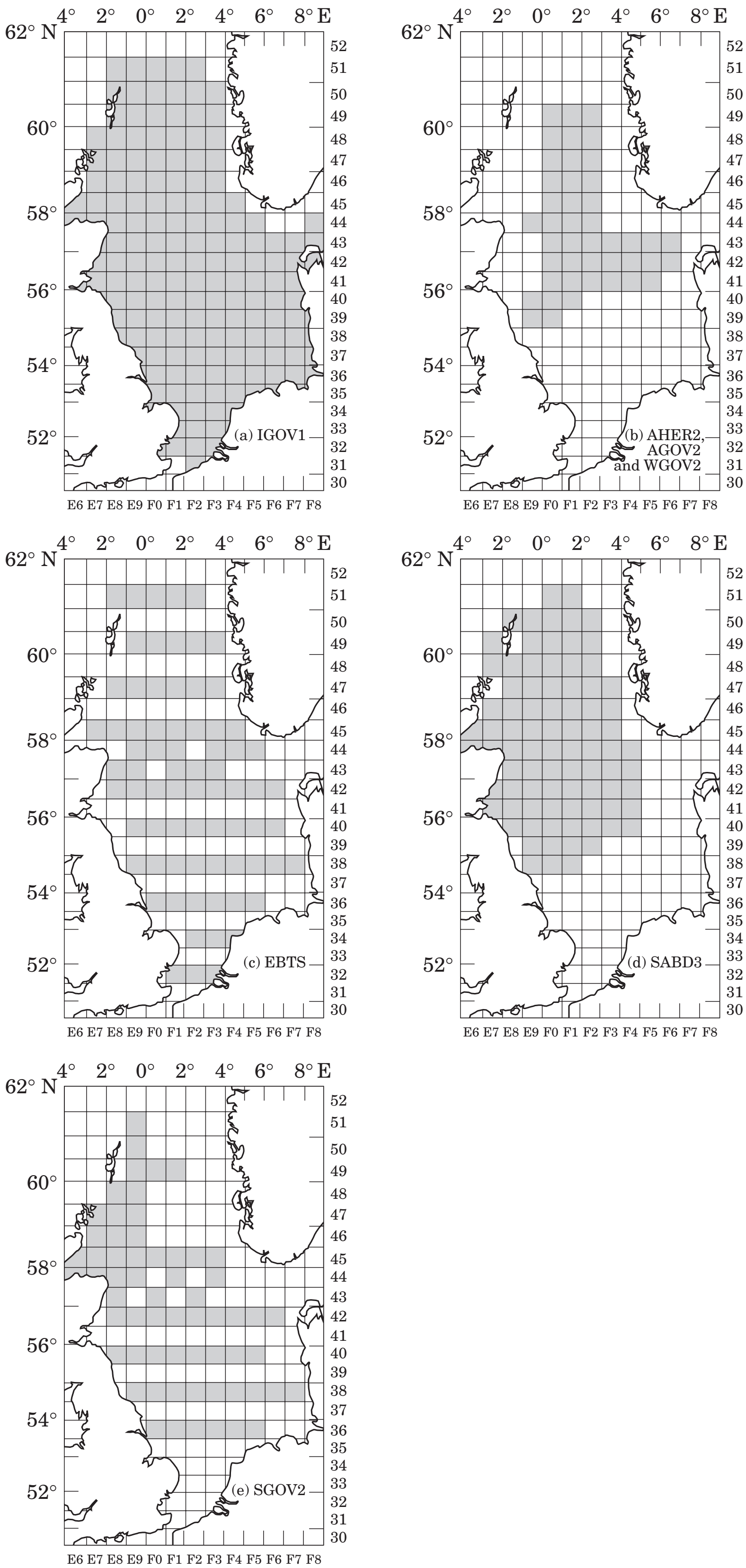

Figure 1. The North Sea showing ICES rectangles and the geographic coverages of each IBTS survey listed in Table 1. (a) IGOV1; (b) DGOV2=same for AHER2, AGOV2, WGOV2; (c) EBTS=same for EGRT3, EGOV3 and EGOV4; (d) SABD3; and (e) SGOV2.

squares. The two variances are independent assuming that the residuals arose from errors within-surveys rather than from a systematically poor fit of the model to certain ages or year classes. The probability of the relative magnitudes of the two variances occurring by chance could therefore be assessed with $F_s = \hat{\sigma}_r^2$. A significant result implies that the assumption $\mathrm{Var}(\varepsilon_{a,c,s}) = \sigma^2$ used for fitting Equation (1) was questionable and that the model might be better fitted with the least precise survey indices removed or down-weighted in some way. However, such operational decisions were not the responsibility of the author. The present study therefore stopped with a detailed examination of the residuals from the significant surveys with the aim of better establishing their relatively poor performance.

The formulae used for the two variances were based on assignation of d.f.:

$$\hat{\sigma}_s^2 = \frac{\sum_a \sum_c \varepsilon_{a,c,s}^2}{n_s - \{(P - k + 1)n_s/n\} - 1} \tag{2a}$$

$$\hat{\sigma}_r^2 = \frac{\sum_a \sum_c \sum_{i \neq s} \varepsilon_{a,c,i}^2}{n - n_s - \{(P - k + 1)(n - n_s)/n\} - k + 2} \tag{2b}$$

Here $n = \Sigma n_s$ is the number of observed abundance indices for all ages from all k surveys, $s = 1, \ldots, k$, and P is the total number of parameters estimated to fit Equation (1). One d.f. was allowed in $\hat{\sigma}_s^2$ for estimation of S_s, and $(k - 2)$ in $\hat{\sigma}_r^2$ for estimation of the other survey intercalibration factors. The remaining $(P - k + 1)$ parameters not associated specifically with a survey accounted for d.f. in the two variances in proportion to the numbers of residuals contributed to each. Note that the sum of the d.f. (denominators) in Equation (2a) and (2b) is $(n - P)$, as usual for the residual variance of a least squares model. The effects on d.f. of correlations between residuals across-ages-within-surveys were ignored at ths stage on the grounds that they would affect the numerator and denominator variances of F_s in approximately similar proportions.

Correcting residual variance for dependent observations

Dependences between residuals-at-age in each annual set of survey indices were assessed by firstly arranging them in a matrix having a colum for each age group in whole years and a row for each annual survey result, and secondly calculating product-moment correlations for each pair of columns. The time-series were not long enough for all surveys to permit reasonable estimates of these correlations separately by individual survey. Auto-correlation of residuals between years for each survey may have been present (Pennington and Godø, 1995) but had to be ignored because of the difficulty of removing it from short series in which error and trend are confounded. Correlations of residuals between surveys within each year were assumed zero because the various vessels of each survey trawled independently at different stations on different dates and at different times of day.

The following estimator of σ^2 was used to allow for age-related dependence among residuals, i.e. for $E(\varepsilon_{a,c,s}\varepsilon_{a',c,s}) \neq 0$ for $a \neq a'$:

$$\hat{\sigma}^2 = \sum_{i=1}^{n} \varepsilon_i^2 / (n - P)Q \tag{3}$$

The d.f. reduction factor, Q, was estimated as summarized in the Appendix. Let A be the number of age-groups in the model. Q may range from 1/A when the residuals-at-age are perfectly dependent across age groups within a year and survey, through to 1 when they are perfectly independent. Standard errors for fitted parameters were computed from $\sqrt{\hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}}$ where $\mathbf{X}$ denotes the predictor matrix for the right hand side of Equation (1), i.e. including all indicator variables for year classes and surveys. The significance of the $\ln(a+1)$ factor in Equation 1 was tested with

$$F = \frac{\text{reduction in sum of squares due to } \ln(a+1)}{\hat{\sigma}^2}$$

with 1 and $(n - P)Q$ d.f. to allow for correlated residuals, the second value being rounded to an integer.

Results

Estimated parameters

The $\ln(a+1)$ factor was significant for all species ($p<0.01$). Estimates of survey intercalibration factors, S_s, as logarithms and as ratios, of Z, and of the coefficient of the $\ln(a+1)$ term are shown in Table 2. The adjusted indices for a representative spread of year classes of cod, haddock, whiting, and Norway pout are shown in Figure 2(a)–(d). It can be seen from these Figures that the indices mostly fell in a uniform band around the fitted year-class curves, also shown in the Figures. Cod residuals [Figure 2(a)] appeared most variable, presumably partly because this was the least abundant species. Residual variability is analysed below. No attempt was made to fit additional parameters to the model because of a reluctance to risk destabilising it, particularly *a posteriori.*

Some of the estimated intercalibration factors as logarithms were not significantly different from zero, i.e. if twice the standαrd error encloses zero, e.g. for SABD3 and cod. Nevertheless, being least squares estimates, they were the best unbiased estimates of relative survey

Table 2. Intercalibration for North Sea IBTS relative to EGOV3 for four species of gadoid: (a) Estimated intercalibration factors as natural logarithms for each survey ± standard errors with (below, in parentheses) their equivalent quotients (survey index/EGOV3 index). (b) Other parameters estimated from Equation (1), residual degrees of freedom before and after correcting for dependences among indices-at-age, and corrected residual variances. Refer to Table 1 for survey abbreviations.

	Cod	Haddock	Whiting	Norway pout
(a) Intercalibration factors				
EGRT3	− 0.49 ± 0.19	− 0.72 ± 0.13	− 0.62 ± 0.14	− 1.57 ± 0.33
	(0.61)	(0.49)	(0.54)	(0.21)
SABD3	− 0.16 ± 0.17	− 0.54 ± 0.12	− 0.55 ± 0.13	− 1.03 ± 0.28
	(0.85)	(0.58)	(0.58)	(0.36)
IGOV1	0.37 ± 0.17	0.30 ± 0.12	0.44 ± 0.13	− 0.13 ± 0.30
	(1.44)	(1.35)	(1.55)	(0.88)
AHER2	− 0.83 ± 0.30	− 1.45 ± 0.22	− 2.08 ± 0.23	—
	(0.44)	(0.23)	(0.12)	
AGOV2	0.12 ± 0.28	− 1.02 ± 0.22	− 1.92 ± 0.22	—
	(1.13)	(0.36)	(0.15)	
WGOV2	0.68 ± 0.19	− 0.08 ± 0.14	− 0.67 ± 0.14	—
	(1.97)	(0.92)	(0.51)	
SGOV2	− 0.06 ± 0.19	− 0.31 ± 0.13	0.41 ± 0.14	—
	(0.94)	(0.73)	(1.51)	
EGOV4	− 0.13 ± 0.19	0.33 ± 0.13	0.40 ± 0.14	− 0.81 ± 0.34
	(0.88)	(1.39)	(1.49)	(0.44)
(b) Other regression data				
Coeff. of total mortality, Z	− 1.34 ± 0.12	− 1.74 ± 0.08	− 2.00 ± 0.09	− 5.21 ± 0.52
Coefficient for ln (Age+1)	1.95 ± 0.55	2.01 ± 0.39	4.48 ± 0.41	10.74 ± 1.90
Uncorrected d.f.	417	412	438	157
Corrected d.f.	347	338	371	147
Corrected $\hat{\sigma}^2$	0.619	0.310	0.381	0.988

effects and were therefore applied, regardless of significance, to adjust their corresponding survey indices to the EGOV3 standard.

There were no clear indications of changes in the slopes of the year-class curves between the 1970s and 1990s, although relatively few data were available in the earlier years. Thus any changes in Z, due, for example to changes in commercial fishing effort, were not detected.

Survey precisions

Analysis of relative residual variances highlighted certain surveys having significantly ($p<0.05$) lower precision for one or two species, given that the model was appropriate. See Table 3. High relative residual variance for cod surveyed by EGRT3 was linked with positive residuals for fish less than three years old and an increased spread of residuals for older fish (Figure 3). The same situation for haddock surveyed by WGOV2 was linked with a peak of residuals in 1989 (Figure 4). Both WGOV2 and EGOV4 stood out as relatively imprecise surveys for whiting, apparently due to trends in the residuals over time in both cases (Figure 5). EGOV3 and, to a lesser extent, EGOV4 were the least precise surveys for Norway pout; residuals were erratic over both time and age [Figure 6(a) and (b)].

Correcting residual variance estimates

Table 4 shows correlation matrices of residuals between age groups 1 to 5 for cod, haddock, and whiting, and 1 to 3 for Norway pout. There were significant ($p<0.05$) positive correlations among most age groups for haddock, and among several for cod and whiting. Standard errors for fitted parameters shown in Table 2 [and the significance of the $\ln(a+1)$ factor in the model, noted above] were therefore estimated after correcting d.f. for the lack of independence between residuals-at-age using Equation (3). Reductions in d.f. and the corresponding increase in $\hat{\sigma}^2$ were 15–20% for cod, haddock, and whiting, and 6% for Norway pout.

Discussion

Statistical estimation of intercalibration factors for abundance indices obtained from different contemporary surveys for the same stock of fish has been demonstrated as a simple exercise in linear modelling when the survival

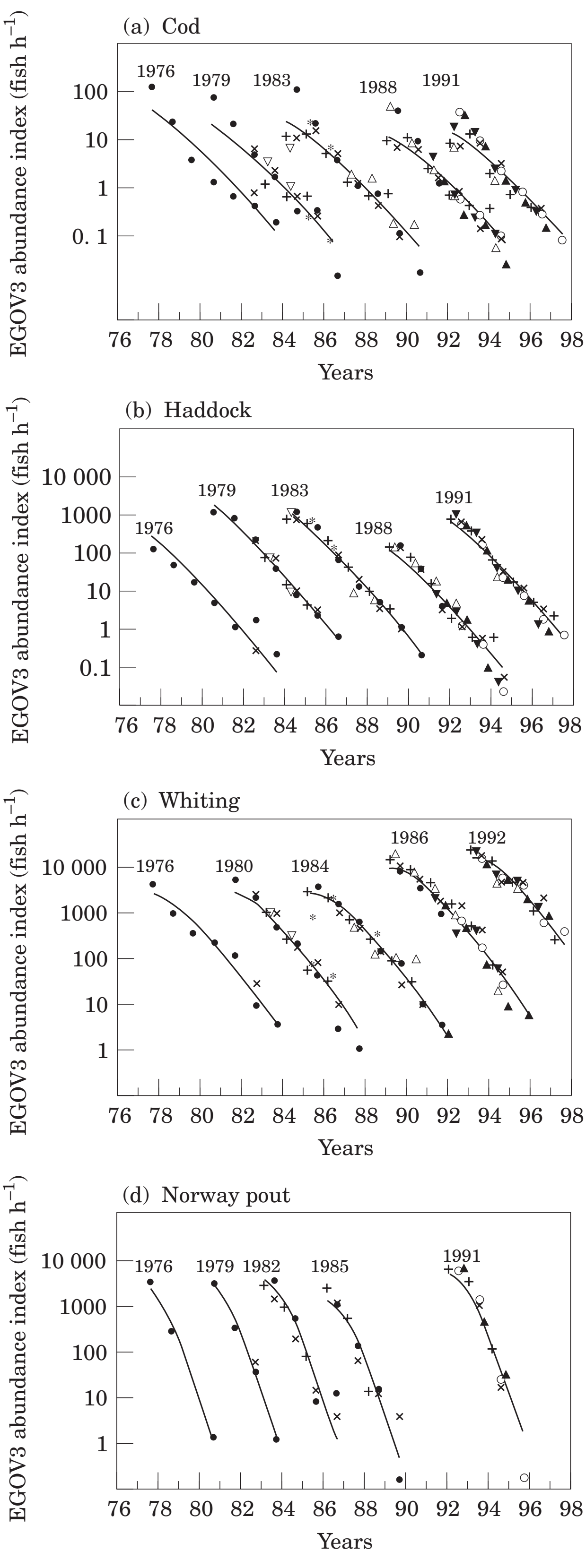

Figure 2. North Sea abundance indices (on a logarithmic scale) for selected year classes after adjusting to an EGOV3 standard, together with fitted year-class curves. (a) Cod; (b) haddock; (c) whiting; (d) Norway pout. Key: EGRT3=dot; SABD3=X; IGOV1=+; AHER2=inverted open triangle; AGOV2=*; WGOV2=open triangle; SGOV2=inverted solid triangle; EGOV4=solid triangle; EGOV3=open circle.

Table 3. Analysis of relative residual variances of North Sea IBTS after fitting year-class curves [Equation (1)] for four species of gadoid. Statistically significant, high values of the F statistics shown indicate high variability of the individual survey relative to all the other surveys. $*=p\le 0.05$, $**=p\le 0.02$, $***=p\le 0.01$; uncorrected degrees of freedom of the single survey and of the remaining surveys (see text) are shown as super- and subscripts on F, respectively. Refer to Table 1 for survey abbreviations.

Survey	Cod	Haddock	Whiting	Norway pout
EGRT3	$F^{95}_{322}=1.48^{***}$	$F^{97}_{315}=1.24$	$F^{97}_{341}=1.18$	$F^{45}_{112}=1.04$
SABD3	$F^{83}_{334}=0.46$	$F^{83}_{329}=0.83$	$F^{84}_{354}=0.54$	$F^{45}_{112}=0.60$
IGOV1	$F^{78}_{339}=1.00$	$F^{83}_{329}=0.71$	$F^{84}_{354}=0.31$	$F^{35}_{122}=0.44$
AHER2	$F^{10}_{407}=0.82$	$F^{8}_{404}=0.40$	$F^{10}_{428}=1.27$	—
AGOV2	$F^{11}_{406}=0.51$	$F^{8}_{404}=0.51$	$F^{10}_{428}=1.68$	—
WGOV2	$F^{40}_{377}=1.16$	$F^{32}_{380}=1.87^{***}$	$F^{44}_{394}=2.39^{***}$	—
SGOV2	$F^{33}_{384}=1.13$	$F^{33}_{379}=0.98$	$F^{33}_{405}=0.52$	—
EGOV4	$F^{33}_{384}=0.95$	$F^{31}_{381}=0.74$	$F^{39}_{399}=1.70^{***}$	$F^{12}_{145}=1.91^{*}$
EGOV3	$F^{33}_{384}=1.31$	$F^{35}_{377}=1.28$	$F^{37}_{401}=1.30$	$F^{20}_{137}=2.55^{***}$

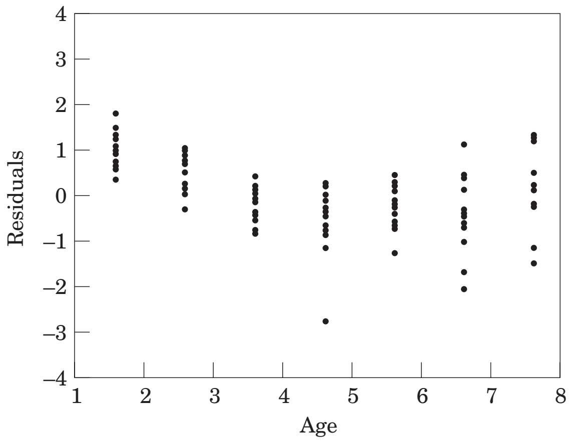

Figure 3. EGRT3 residuals for cod after fitting Equation (1) showing pattern over age.

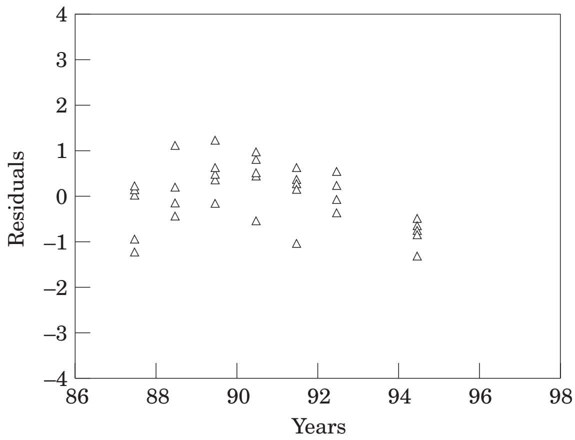

Figure 4. WGOV2 residuals for cod after fitting Equation (1) showing pattern over time.

of year classes can be tracked over periods of several years. Estimates for Norway pout were least precise, presumably because fewer survey results were available and because this species has a relatively short lifespan giving fewer data points for estimating the slopes of the year-class curves. In favourable circumstances, statistical modelling of abundance indices can be considered as a useful alternative to comparative trawling experiments. The resulting intercalibration factors combine the net effects on abundance indices of all the specific spatial, temporal, ship- and gear-related differences between two surveys. If only one effect is different, as between EGRT3 and EGOV3 (gear only), or between EGOV3 and EGOV4 (season only), see Table 1, the intercalibration factor is an estimate of that effect. Otherwise, confounding prevents estimation of the individual effects. By contrast, comparative trawling trials can be designed as experiments to estimate specific

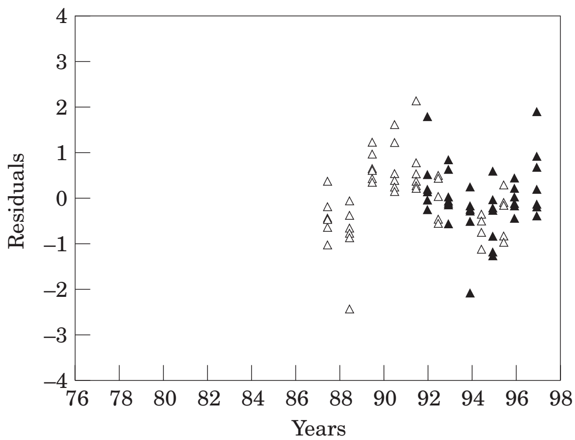

Figure 5. WGOV2 and EGOV4 residuals for whiting after fitting Equation (1) showing patterns over time. See Figure 2 for key to symbols.

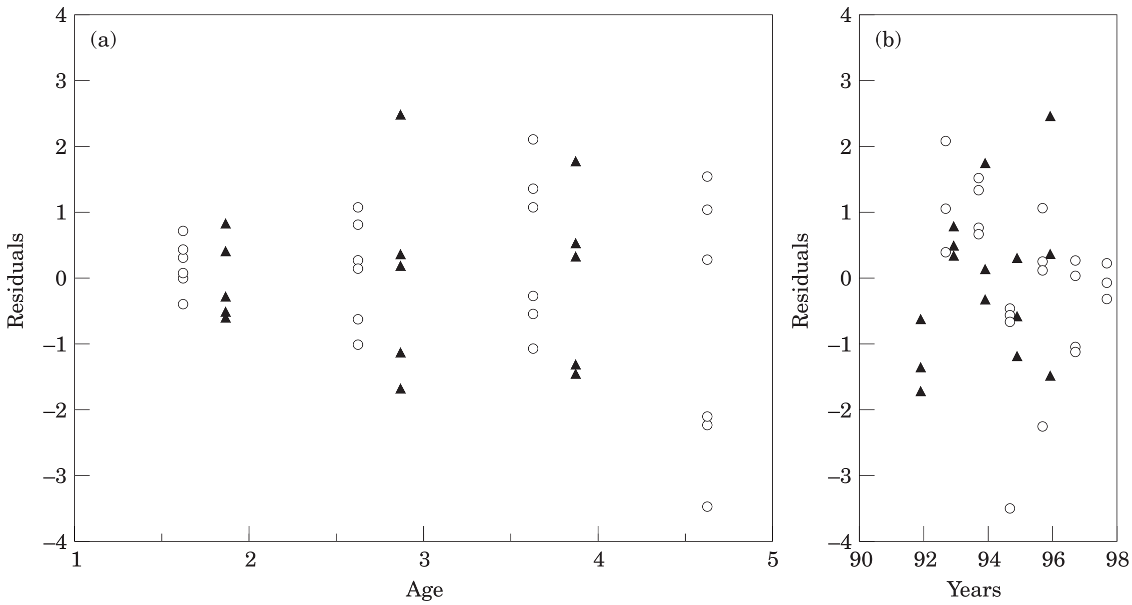

Figure 6. EGOV3 and EGOV4 residuals for Norway pout after fitting Equation (1) showing patterns over (a) time and (b) age. See Figure 2 for key to symbols.

Table 4. Correlation analysis of residuals of North Sea IBTS after fitting year-class curves [Equation (1)] for four species of gadoid. *=p<0.05; **=p<0.01.

Age	2	3	4	5
(a) Cod				
1	0.37**	− 0.02	− 0.34*	− 0.31*
2		0.13	− 0.16	− 0.23
3			0.32*	0.19
4				0.27
(b) Haddock				
1	0.58**	0.42**	0.29*	0.06
2		0.54**	0.33*	0.06
3			0.35*	0.11
4				0.12
(c) Whiting				
1	0.52**	0.11	0.10	0.07
2		0.60**	0.40**	0.21
3			0.53**	0.50**
4				0.42**
(d) Norway pout				
1	0.29	− 0.09		
2		0.21		

effects at the locality and season of the trials. However, applying those estimates to intercalibrate whole-survey abundance indices involves the assumption that the effects can be extrapolated to the whole survey region and season. This could be a major weakness, depending on the extent and duration of the trials.

An analysis of relative residual variances in a group of surveys is bound to find a certain ordering, with one less precise than all the others. Also, although differences in variance may be quite small in practical terms, statistical significance may arise because, as in the present case, large numbers of observations confer high power on the F test. Nevertheless, surveys with significantly high relative residual variance deserve careful scrutiny. Patterns of residual variability over time and age (Figures 3–6) are likely to suggest reasons for relative imprecision. They may be biological, e.g. year-to-year variation in

- the match between the survey area and the area occupied by the stock at the time of the survey (Swain and Sinclair, 1994); or in
- the growth rates of young fish which influence when they become catchable by the survey trawl (Godø and Sunnanå, 1992).

Alternatively, the reasons may be procedural, related to e.g.

- the methods of sampling catches on deck, reading otoliths, and formulation of age-length keys; and
- trawling technique and gear geometry.

Application of the Q factor in Equation (3) to correct d.f. and to inflate estimates of residual variances produced corrections which were 20% or less. The small size of these corrections may partly be a result of the widespread practice of sampling catches in order to reduce the numbers of fish measured and aged to manageable quantities. This superimposes negative covariance on the positive covariance among numbers-at-age arising from trawling (Cotter, 1998), implying a degree of cancellation. Measurement noise would also mask positive dependences between estimated abundances at different ages.

Estimates of Z from the present study of IBTS data (Table 2) may be compared with those of previous investigators. Jensen (1939) collated published data to prepare stock curves annotated with annual percentage decreases, d, for several species and seas. These may be converted using $Z = \ln\{(100-d)/100\}$. For North Sea cod in the 1930s, Jensen's data give $Z = -1.08$ between 2 and 5 years old, and -0.54 for older fish. Comparing these estimates with $Z = -1.33$ in Table 2 suggests that an increase in fishing effort may have been responsible for the difference. However, larger absolute values to -1.71 were reported by Jensen for cod towards and into the Baltic at about the same time. Possibly, migration of cod from the Baltic to the North Sea was occurring. For North Sea haddock, Jensen demonstrated that Z increased in magnitude from -0.82 to -1.31 corresponding to the introduction of the Vigneron–Dahl trawl in the late 1920s. This, and the value -1.74 for haddock in Table 2 suggest that increases in fishing effort since the 1920s have been detected.

Cook (1997) used IBTS data to estimate fishing mortality, F, for North Sea fish over 3 years old. Noting that $Z = F + M$, his estimates can be compared approximately with Z in Table 2 by subtracting $M = 0.2$ from the Z. This is a standard value used for stock assessments; larger M are used for younger fish (ICES, 1998b). Firstly, Cook reported generally flat trends in F from 1982 for cod, haddock and whiting in agreement with my finding of constant Zs for these species. The main commercial fishing method for these species in the North Sea is the otter trawl. According to Jennings *et al.* (1999), total otter trawling effort by England, Germany, Norway, Scotland and Wales declined from approximately 1.2×10^6 h in 1978 to 1.0×10^6 h in 1995. This relatively small change, i.e. 20%, suggest that the failure to find trends in F over the period was not serious, especially as the efficiency of trawling probably increased over the period, counter-acting the effect of the decline in fishing hours. The estimates of Z in Table 2 for cod, haddock, and whiting are of somewhat greater magnitude than those shown as time-series by Cook (1997). This might be explained by our different modelling approaches, but my estimate of -5.21 for Z for Norway pout differs substantially from Working Group estimates ranging from -1.2 in 1985 down to -0.8 in the 1990s, given $M = 0.4$ for all ages (ICES, 1998b). This suggests either that the slopes of the year-class curves for Norway pout estimated here were seriously biased or that the Working Group's estimate of Z was too low.

Acknowledgements

I am grateful to T. Boon, and J. Dann of CEFAS, Lowestoft, for advice on the English surveys, to A. Newton of the Fisheries Research Services Marine Laboratory, Aberdeen for advice on Scottish surveys, and to S. Ehrich of the Institut für Seefischerei, Hamburg for advice on German surveys. They are, however, not responsible for any mistaken information about the surveys in this paper. M. Bravington, CEFAS Lowestoft, helpfully commented on a draft manuscript, as did referees of this journal. The research was funded by the Ministry of Agriculture, Fisheries and Food, London, but no statements in this paper should be construed as official policy.

References

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

Cotter, A. J. R. 1993. Intercalibration of groundfish surveys using regression analysis of yearclass mortalities. ICES CM 1993/G: 21.

Cotter, A. J. R. 1994. Compliance testing of two or more water quality determinands using quantiles. Environmetrics, 5: 29–45.

Cotter, A. J. R. 1998. Method for estimating variability due to sampling of catches on a trawl survey. Canadian Journal of Fisheries and Aquatic Sciences, 55: 1607–1617.

Ehrich, S. 1991. Comparative fishing experiments by research trawlers for cod and haddock in the North Sea. Journal du Conseil Internatinal pour l'Exploration de la Mer, 45: 275–283.

Godø, O. R. 1994. Factors affecting the reliability of groundfish abundance estimates from bottom trawl surveys. *In* Marine fish behaviour in capture and abundance estimation, pp. 166–199. Ed. by A. Ferno, and S. Olsen. Fishing News Books, Farnham, Surrey, England.

Godø, O. R., and Sunnanå, K. 1992. Size selection during trawl sampling of cod and haddock and its effect on abundance indices at age. Fisheries Research, 13: 293–310.

Heessen, H. J. L., Dalskov, J., and Cook, R. M. 1997. The international bottom trawl survey in the North Sea, the Skagerrak and Kattegat. ICES CM 1997/Y: 31.

ICES 1992. Report of the workshop on the analysis of trawl survey data, Woods Hole. ICES CM 1992/D: 6.

ICES 1996a. Report of the international bottom trawl survey working group; Addendum: Manual for the international bottom trawl surveys. Revision V. ICES CM 1996/H: 1.

ICES 1996b. Report of the working group on the assessment of demersal stocks in the North Sea and Skagerrak. ICES CM 1996/Assess: 6.

ICES 1998a. Report of the study group on the evaluation of the quarterly IBTS surveys. ICES CM 1998/D: 4.

ICES 1998b. Report of the working group on the assessment of demersal stocks in the North Sea and Skagerrak, October 1997. ICES CM 1998/Assess: 7.

Jennings, S., Alvsvåg, J., Cotter, A. J. R., Ehrich, S., Greenstreet, S. P. R., Jarre-Teichmann, A., Mergardt, N., Rijnsdorp, A. D., and Smedstad, O. 1999. Fishing effects in northeast Atlantic shelf seas: patterns in fishing effort, diversity and community structure. III. International trawling effort in the North Sea: an analysis of spatial and temporal trends. Fisheries Research, 40: 125–134.

Jensen, A. J. C. 1939. On the laws of decrease in fish stocks. Rapports et Procès-Verbaux des Réunion du Conseil International pour l'Exploration de la Mer, 110: 85–96.

King, M. 1995. Fisheries Biology, Assessment and Management. Fishing News Books, Blackwell Science Ltd, Oxford. 341 pp.

Mitson, R. B. 1995. Underwater noise of research vessels – review and recommendations. ICES Cooperative Research Report 209, 61 pp.

Munro, P. T. 1998. A decision rule based on the mean square error for correcting relative fishing power differences in trawl survey data. Fishery Bulletin, 96: 538–546.

Myers, R. A., and Cadigan, N. G. 1995. Statistical analysis of catch-at-age data with correlated errors. Canadian Journal of Fisheries and Aquatic Sciences, 52: 1265–1273.

Myers, R. A., and Stokes, K. 1989. Density-dependent habitat utilization of groundfish and the improvement of research surveys. ICES CM 1989/D: 15.

Ona, E., and Godø, O. R. 1990. Fish reaction to trawling noise: the significance for trawl sampling. Rapports et Procès-Verbaux des Réunion du Conseil International pour l'Exploration de la Mer, 189: 159–166.

Paloheimo, J. E., and Dickie, L. M. 1964. Abundance and fishing success. Rapports et Procès-Verbaux des Réunion du Conseil International pour l'Exploration de la Mer, 155: 152–163.

Pelletier, D. 1998. Intercalibration of research survey vessels in fisheries: a review and an application. Canadian Journal of Fisheries and Aquatic Sciences, 55: 2672–2690.

Pennington, M., and Godø, O. R. 1995. Measuring the effect of changes in catchability on the variance of marine survey abundance indices. Fisheries Research, 23: 301–310.

Pennington, M., and Vølstad, J. H. 1994. Assessing the effects of intra-haul correlation and variable density on estimates of population characteristics from marine surveys. Biometrics, 50: 725–732.

Ricker, W. E. 1975. Computation and interpretation of biological statistics of fish populations. Bulletin of the Fisheries Research Board of Canada, Vol. 191, 382 pp.

Sparholt, H. 1990. Using GLM analysis on the IYFS herring data for the North Sea. ICES CM 1990/H: 6.

Swain, D. P., and Sinclair, A. F. 1994. Fish distribution and catchability: what is the appropriate measure of distribution? Canadian Journal of Fisheries and Aquatic Sciences, 51: 1046–1054.

Wilderbuer, T. K., Kappenman, R. F., and Gunderson, D. R. 1998. Analysis of fishing power correction factor estimates from a trawl comparison experiment. North American Journal of Fisheries Management, 18: 11–18.

Appendix

Calculation of Q for Equation (3)

The residuals from fitting Equation (1), $\varepsilon_{a,c,s}$, were arranged into a matrix, **E,** having one column for each age group in integer years and one row for each annual survey result. Ideally, all age groups used in the fit would have been represented but missing values for some years and surveys in the older age columns necessitated deletion of the entire column making the correction to degrees of freedom more approximate. Given A columns in the final matrix with correlations within rows but independence between them, an equivalent number of independent notional variable columns, f, can be estimated from an information statistic, I, described by Cotter (1994). The median, $\tilde{\mu}$ of all elements in **E** was applied to form a matching matrix **M** in which each element

$$m_{a,c,s}=\begin{cases}0 & \text{if } \varepsilon_{a,c,s}\le\tilde{\mu}\\ 1 & \text{otherwise}\end{cases}$$

[Note that $\tilde{\mu}\approx 0$, given $E(\varepsilon_{a,c,s})=0$ and a symmetric distribution of residuals.] Each row in **M** forms a binary number in the range $b=0, \ldots, 2^A-1$. Let there be n_b rows having the same binary value b. Then I was estimated as

$$\hat{I}=-N^{-1}\Sigma_b n_b \ln(n_b/N)$$

and

$$f=\hat{I}/(-\ln 0.5)$$

the divisor representing the information content of one variable around its median. Lastly, $Q=f/A$.