Evaluation of the quality of the North Sea herring assessment

E. John Simmonds


The assessment of North Sea herring has been used to give advice on catch quota for more than 20 years. The data sources comprise acoustic surveys, International Bottom Trawl Surveys, Methot Isaacs–Kidd net post-larval surveys, larval surveys, and catch-at-age data. These sources and their uses are briefly reviewed, and the changes in the weighting attached to each index in the assessment over time are discussed. The performance of the assessment is examined both in historical and analytical retrospectives of spawning–stock biomass and fishing mortality, and in retrospective assessments of numbers by cohort. Increased length of the time-series, the use of a statistical model with appropriate weighting, and a more consistent management strategy have all contributed to the assessment becoming highly stable from one year to the next. The results presented lead to the conclusion that the assessments provide an excellent basis for the management of this stock.

Keywords: North Sea herring, stock assessment, stock management.

Received 29 August 2008; accepted 21 January 2009; advance access publication 22 April 2009.

Introduction

For several hundreds of years, herring (Clupea harengus) have been an important source of food and economic wealth in Europe, with the Baltic and North Sea (NS) resources contributing to the development of the Hanseatic League during the 14th–16th centuries, and conflicting interests in the NS herring fishery being one of the causes of the war between the UK and the Netherlands in the 16th century (Tracy, 1993). Historical detail for the period 1600–1860 is provided by Poulsen (2006), who concludes that, despite periods of hardship, the fisheries then would have had a negligible impact on the stock. By the 1950s, however, the NS herring fisheries had expanded to a level where they were seen to have a major impact on the stock (Simmonds, 2007). Since 1960, the stock has experienced three periods of decline, the first two as a result of severe exploitation. During the first period (1960–1980), international data collection was poorly coordinated and an analytical assessment was not possible. Although scientists provided recommendations for a reduction in the fishery, this advice was acted on too late (Simmonds, 2007). Early in this period, a number of surveys that have subsequently formed the basis for assessment and management decisions was started or reorganized. In reviewing management actions from 1964 to 1978, Saville and Bailey (1980) concluded that the fisheries had been the major cause of extreme stock depletion. Bjornadal and Conrad (1987) went even further and suggested that the stock would have become extinct if the fishery had not been closed in 1978. Reviewing the second decline and subsequent recovery (1990–2002), Simmonds (2007) concluded that data collection and improved assessment techniques had contributed substantially to the adoption of the appropriate management actions that led to subsequent stock recovery.

From 2002 to 2007, recruitment to the stock has been at ~40% of the long-term mean (1984–2001), which has resulted in a third decline. This time, it is not excess fishing mortality per se that has caused decline but reduced recruitment (Payne et al., 2009). The management response in the 1990s, the expansion of the fishery in 2000, and the subsequent decline in total allowable catches (TACs) from 2005 to 2008 have all been based on management advice coming from the ICES stock assessment. I examine the quality of the data entering the NS herring stock assessment and evaluate its performance in relation to the provision of management advice. NS herring is predominantly an autumn- and early winter-spawning herring; larvae develop over winter and may metamorphose either in the following year or in the same year, depending on whether they hatch before or after 1 January. These herring, in common with many fish, are aged by counting rings in the otoliths, and they acquire their first ring in their second winter. Throughout this paper, we use age 0 until 1 January that follows the autumn of hatching age is incremented by 1 at that point, and again on each subsequent 1 January. Care should be taken when comparing this study with other work on autumn-spawning herring in which age is expressed in winter rings. Winter-ring age may be converted to true age by adding 1.

Data used

The assessment uses a variety of data derived from commercial catches as well as fishery-independent trawl, acoustic, and larval surveys. The following provides a historical view of the data, as well as some insight into its extent and quality.

Catch data

Individual nations collect fishery data from landings and at-sea sampling. While targets have been set for numbers of samples (EC, 2001), there is no overall coordination of data-collection methods. Each nation organizes its own scheme to provide national estimates of total landings and their catch-at-age.
composition. An evaluation of the sampling precision in 1998–2000 (ICES, 2001; Simmonds, 2003a) provided estimates of variability in the numbers-at-age landed as used in the assessment. Typically, the precision is highest for age 3, with a coefficient of variation of around 7%, rising to 30% at age 9+ and to 15% at age 0.

Survey data
Sampling of herring larvae started in 1946 (Heath, 1993). These efforts became ICES-coordinated in 1967 (Saville, 1970), and by 1971, a fully coordinated larval survey was initiated across all major spawning areas. Survey effort increased in the 1970s, maintained full coverage in the 1980s, and eventually covered 11 separate regional and seasonal spawning concentrations along the western side of the NS (Heath, 1993). Survey effort declined again in the 1990s. The use of this survey as an index of spawning-stock biomass (SSB) was proposed by Corten (1980). Burd (1985) and Santiago (1986) proposed single-stock virtual population analysis (VPA) that made use of an SSB index based on the integrated larval abundance over all areas sampled. The single-index approach was not implemented at the time, and management was still based on split stocks. It was not until 1989 that the survey was used in a regression with VPA to give management advice. However, the substantial effort spent on this survey in terms of vessel time (exceeding 200 vessel days per year) could not be maintained, and sampling intensity declined soon after, despite favourable reviews by Heath (1993). To accommodate this change, a modelled larval index (MLAI), based on a multiplicative season and area model, has provided an SSB index that goes back to 1973 (Patterson and Beveridge, 1995; Rohlf et al., 1998; Gröger and Schnack 1999; Gröger et al., 2000). The survey is still carried out, using around 80 vessel days (ICES, 2008a).

The ICES Young Herring Survey was initiated in 1962. Subsequently, a series of independent bottom-trawl surveys, carried out in the first quarter of the year, was brought together under the auspices of ICES, which, by the early 1970s, covered the whole NS. ICES (1975) made proposals for standardizing gear and sampling protocols, and subsequently this survey became fully coordinated. By 1984, all vessels were using the same gear and applied standard procedures, and the survey became the International Bottom Trawl Survey (IBTS). From 1991 to 1996, the survey was carried out on a quarterly basis and, since 1997, on a half-yearly basis (first and third quarter). However, only the first-quarter survey extends back to the early 1970s. The survey provides abundance estimates for ages 2–6 herring (ICES, 2008b), and data from 1984 on are used in the assessment.

In 1977, the Methot Isaacs–Kidd (MIK) net survey was initiated as an integral part of data collection during the first-quarter IBTS (Corten and Van der Kamp, 1981; Munk, 1986; Munk and Christensen, 1990). This ichthyoplankton survey is carried out at night, whereas all trawls have to be made in daylight. Therefore, it can be included in the operations with minimal extra cost. By 1992, procedures and methods had also been standardized across the fleet of participating vessels. The MIK survey provides an estimate of surviving post-larvae (age 1) from the autumn spawning of the previous year, and the index proves to be a good estimator of recruitment (ICES, 2008b). Data from 1992 on are used in the assessment.

In 1979, an acoustic survey was attempted (ICES, 1980). Initially, the results lacked coherency, although they seemed promising. Standardization appeared to present a problem and, in 1981, ICES-coordinated work commenced on the development of standard calibration procedures. By 1985, the survey was able to provide coherent results among the participating vessels (Simmonds et al., 1986). These efforts subsequently resulted in the standard acoustic-survey calibration manual (Foote et al., 1987). It took until 1989 before the timing and area of coverage were standardized and the acoustic survey started to yield a reliable time-series of estimates of numbers-at-age in the stock, stock weights, and fraction mature-at-age (ICES, 2008a).

The coherence in the various abundance estimates from surveys may be visualized in normalized scatterplots showing how estimates of numbers-at-age correspond to the same cohort estimated 1 year later as numbers-at-age+1 within a survey, or how they correspond between the same age groups (or SSB indices) among surveys (Figure 1). The correlation matrix (Table 1) represents a basic metric for information content (or coherence) in the data. All correlations are significant, except for two involving the IBTS index at age 3. A comparison of the correlation coefficients from different series is only considered broadly indicative of the differences in performance, because although they are usually significant, their magnitudes depend on the range of recruitment values in the series. These ranges differ slightly because the periods covered by each survey and age class are different. The acoustic survey is the most self-consistent dataset. The acoustic estimate for age 2 is the most consistent with the age-1 index from the MIK survey, showing only slightly better agreement than the IBTS age-2 index. However, the IBTS age-2 index still provides an important contribution to the assessment. The IBTS indices are less consistent from age 3 on. Interestingly, the correlations for age 3 seem worse than for some of the older ages; also the acoustic survey shows greater consistency at mid-ages 4–6 than between ages 2 and 3, although sampling error on both surveys is less at age 3 than for older ages (Simmonds, 2003b). This suggests that some of the variability in availability to the survey gear may be due to different behaviour among age groups as they partially mature. Therefore, it may be better to estimate year classes at points of stability (fully immature or mature) than during periods of flux.

A comparison between the MLAI and the SSB derived from the acoustic survey suggests a non-linear relationship. Although the acoustic survey is not fitted explicitly as an SSB index in the assessment model, the numbers-at-age, mean weights-at-age, and fraction mature-at-age come from this survey, so a fit is implied. To accommodate the apparent curvature in the relationship between the SSB and the MLAI, the relationship is fitted in the assessment with a non-linear model.

Assessments
Historical changes
Before the early 1990s, NS herring was assessed as two separate stocks (northern NS and Downs), and the youngest ages (ages 1 and 2) found in the Skagerrak and Kattegat were often excluded from the analysis because they could not be assigned to either stock. As these stock units did not correspond to the current definition of management areas, the older ICES reports cannot be used to derive comparable assessments of the single NS stock complex. In 1991, VPA estimates of adult northern
(autumn-spawning) and southern (Downs winter-spawning) components were provided separately, and the juvenile fisheries were explicitly included in the VPA. By 1993, problems caused by keeping separate the mixed components taken by fisheries in different areas and seasons had been recognized, and a single-area assessment was established. From 1991 to 1995, separable VPA was used to estimate the stock, either by single unit or in separate areas, but retrospective performance was poor. From 1996 on the advice has been based on an integrated catch analysis (ICA) model (Patterson and Melvin, 1996).

Variance weighting factors
Initially (up to 2002), all catch-at-age data used in the ICA assessment were treated equally, whereas survey estimates of abundance for multiple ages from the same survey were assumed to be highly correlated and downweighted to $1/n$, where $n$ is the number of age classes included in the survey index. After investigating sampling precision of input data, the weighting of data in the model was reviewed and adjusted. Simmonds (2003b) reported on the methods by which the new weighting factors were derived. The use of the variance of the natural logarithm of the observations was based on a two-stage bootstrap procedure. This choice matches the maximum log-likelihood method with a lognormal error distribution used within the ICA model (Patterson and Melvin, 1996). All indices were treated in the same manner. The “block” estimates-at-age are bootstrapped using a resampling procedure with replacement (Efron and Tibshirani, 1993). For trawl surveys, a block is represented by each trawl station, whereas for acoustic surveys, it is the mean of track segments, effectively within an ICES rectangle. The use of different degrees of spatial averaging (supports) for trawl and acoustic surveys is inevitable given the different sampling methodology, but the overall variance is not sensitive to this choice of support (see section 4.3 in Rivoirard et al., 2000). The bootstrap analysis provides a variance/covariance estimate for each index-at-age, based on an assumption of identically but independently distributed (iid) samples. However, as the spatial distributions are correlated and survey sampling is non-random in space, the spatial autocorrelation has been taken into account using geostatistics. This part of the methodology is described in Rivoirard et al. (2000), who provide the formulae and methods required to estimate variograms and calculate the estimation variance. Petitgas and Lafont (1997) provide software (EVA2) that has been used here for calculating the estimation variance for all survey indices. The iid estimates are corrected to provide overall estimates of variance/covariance across ages for each survey.

The mean variance/covariance estimates for the time-series were calculated to provide one average variance/covariance matrix per survey. The ICA model does not deal explicitly with covariance in input data (in common with many simple assessment models), but it does allow modification of weighting factors-at-age ($w_{age}$) to account for this in a general way. In the fit, the weights are applied to the squared residuals, which is analogous to an inverse variance-weighting method. ICA provides two methods for applying weights: (i) prior weighting and (ii) adaptive weighting (based on inverse variance of residuals in a survey series). Here, the prior weighting is used, based on the sampling precision described above. To account for covariance between estimates at adjacent ages in the same survey, the weighting factors are reduced in the following way. First consider the two extreme cases: (i) zero correlation—the weights used would be unmodified; (ii) the correlation of 1 over $n$ ages—the weights would be reduced by a factor of $1/n$, because there would be no independence between estimates at each age. For both the IBTS and the acoustic survey, correlation between adjacent ages lies between these two extremes of 0 and 1. For both surveys, the estimates of abundance at the youngest and subsequent ages (ages 2 and 3) appear to be uncorrelated, and there is no confounding ageing error at these ages because their length distributions show no overlap. Therefore, the lowest age is given a covariance weighting of unity, and the sampling variance weight is unaltered. For all
Table 1. Correlation coefficients within and between surveys.

<table>
<thead>
<tr>
<th>Age</th>
<th>MLAI-AS</th>
<th>MIK-IBTS</th>
<th>MIK-AS</th>
<th>IBTS-IBTS</th>
<th>AS-AS</th>
<th>IBTS-AS</th>
<th>AS-IBTS</th>
<th>IBTS-AS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 19)</td>
<td>(n = 14)</td>
<td>(n = 13)</td>
<td>(n = 29)</td>
<td>(n = 18)</td>
<td>(n = 18)</td>
<td>(n = 18)</td>
<td>(n = 19)</td>
</tr>
<tr>
<td>1–2</td>
<td>–</td>
<td>0.70</td>
<td>0.86</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>2–3</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.46</td>
<td>0.91</td>
<td>0.58</td>
<td>0.57</td>
<td>Age 2–2</td>
</tr>
<tr>
<td>3–4</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>(0.13)</td>
<td>0.92</td>
<td>(0.28)</td>
<td>0.62</td>
<td>Age 3–3</td>
</tr>
<tr>
<td>4–5</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.33</td>
<td>0.95</td>
<td>0.57</td>
<td>0.76</td>
<td>Age 4–4</td>
</tr>
<tr>
<td>5–6</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.69</td>
<td>0.96</td>
<td>0.82</td>
<td>0.78</td>
<td>Age 5–5</td>
</tr>
<tr>
<td>6–7</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.88</td>
<td>0.88</td>
<td>0.82</td>
<td>–</td>
<td>Age 6–6</td>
</tr>
<tr>
<td>7–8</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.93</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>8–9</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.28</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>SSB</td>
<td>0.70</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

AS, acoustic survey; n, number of paired observations; values in parentheses, not significant under the assumption of positive correlation only; IBTS, International Bottom Trawl Survey; MIK, Methot Isaacs–Kidd; MLAI, modelled larval index; SSB, spawning-stock biomass.

older ages, there is evidence of correlation in the estimates, both in the sampling and because of potential ageing errors; overall, this covariance is indicated by positive values next to the diagonal in the covariance matrix (not shown). The weighting for older ages has been modified to account for the observed values of covariance to give \( \lambda_{\text{age}} \) model according to

\[
\lambda_{\text{age}} = \frac{1}{\text{var}_{\text{age}} \cdot \text{cov}_{\text{age,age-1}} / \sum_i \text{cov}_{\text{age,age-1}}},
\]

where \( \text{var}_{\text{age}} \) is the variance of ln(estimate-at-age), \( \text{cov}_{\text{age,age-1}} \) the covariance(age, age − 1), and \( n \) the number of ages in the correlated sequence.

The resulting weighting factors, used since 2006, are given in Table 2. These differ only slightly from the original factors, used first in the 2002 assessment. One key aspect is that this choice of prior weighting leads to weighting factors that become smaller first in the 2002 assessment. One key aspect is that this choice of adaptive weighting available in ICA and in other assessment models, such as XSA, tends to give higher weights to older ages. Part of the reason is that these types of VPA model can fit flexibly to extreme (oldest or youngest) ages while having greater difficulty in fitting to mid-ages (similar to the sensitivity of linear regression models to values at the extremes of the data range). Therefore, adaptive models are apt to overemphasize information from older age groups that are potentially more noisy. This does also happen if ICA is used in this manner (ICES, 2001) and would lead to poorer retrospective performance. Using a priori weights based on precision reduces this problem and contributes to a more reliable assessment. The detailed formulation of the final model is given in the Appendix.

Table 2. Weighting factors based on bootstrap estimates of survey error and catch sampling error after correction for spatial autocorrelation in the surveys and correlation between estimates at adjacent ages (values scaled to an arbitrary mean because only the relative value influences the assessment).

<table>
<thead>
<tr>
<th>Age</th>
<th>Catch (( \lambda_{\text{catch}} ))</th>
<th>AS (( \lambda_{\text{AS}} ))</th>
<th>IBTS (( \lambda_{\text{IBTS}} ))</th>
<th>MIK (( \lambda_{\text{MIK}} ))</th>
<th>MLAI (( \lambda_{\text{MLAI}} ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.10</td>
<td>0.63</td>
<td>0.47</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>2</td>
<td>0.10</td>
<td>0.63</td>
<td>0.47</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>3</td>
<td>3.67</td>
<td>0.62</td>
<td>0.28</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>4</td>
<td>2.87</td>
<td>0.17</td>
<td>0.01</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>5</td>
<td>2.23</td>
<td>0.10</td>
<td>0.01</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>6</td>
<td>1.74</td>
<td>0.09</td>
<td>0.01</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>7</td>
<td>1.37</td>
<td>0.08</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>8</td>
<td>1.04</td>
<td>0.07</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>9</td>
<td>0.94</td>
<td>0.07</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>10+</td>
<td>0.91</td>
<td>0.05</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>SSB</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.60</td>
</tr>
</tbody>
</table>

AS, acoustic survey; IBTS, International Bottom Trawl Survey; MIK, Methot Isaacs–Kidd; MLAI, modelled larval index; SSB, spawning-stock biomass.

Analysis of historical assessment performance

Two types of analysis were carried out: (i) a “historical retrospective”, evaluating the consistency between the estimates of relevant parameters published by ICES in consecutive years, and (ii) an “analytical retrospective”, using current data and current assessment methods on progressively truncated time-series.

For the historical retrospective, the ICES assessment quality-control output dates back to 1996. This set was extended back to 1991 by extracting data from the reports of the annual ICES Working Groups and the Advisory Committee on Fisheries Management. For 1994 and 1995, this was straightforward because NS herring were assessed as a single stock. For 1991–1993, NS herring were assessed as two stocks and the assessments had to be combined to produce comparable, although artificial, “historical quality” estimates. Before 1991, the available information was incomplete, so a comparison was not possible.

For the analytical retrospective, assessments for truncated time-series could be extended back to 1998 on the same basis as the current assessment. For earlier years, it is not possible to maintain the same analysis without changing some datasets, because the age-2 index from the acoustic survey is not available before 1995 and the MIK time-series has been truncated to 1992. To extend a coherent analysis back to 1992, the dataset was reconstructed by: (i) removing the age-2 index from the acoustic survey time-series (1995–2007) to give a coherent set dating back to its start in 1989, and (ii) extending the MIK dataset back to 1989 (thus matching the length of the acoustic dataset). Estimates of MIK are available from 1977 to 1991, but the quality is questionable, especially for the earlier surveys. However, including the values for 1989–1991 did not change the survey catchability (\( q \)), and they therefore appear to be unbiased. In contrast, the use of a longer series of data (back to 1977) is more problematic because this does change the \( q \). As the inclusion of potentially biased data might adversely affect the results of the analysis and, in any case, would not have allowed an analytical retrospective before 1991, MIK values were limited to those back to 1989.
Figure 2. Historical retrospective: sequential assessment estimates of: (a) SSB, (b) average fishing mortality of ages 2–6 for the years 1991–2007, and (c) recruitment (numbers at age 0) by year class in assessment years 1992–2008 (ICES quality-control output, adjusted for earlier years to account for differences in the unit stock assessed).
Analytical retrospectives for this reconstructed dataset were carried out using the ICA weighting model (the Appendix) with the current weighting factors (Table 2).

To evaluate the performance of the assessment in terms of consistency of annual advice, the chosen metric is the relative change in estimated SSB, fishing mortality (F), and recruitment at age 1 (R) over time. This relative change is estimated as \( 1 - \frac{A_{y+1}}{A_y} \), where \( A_y \) is the estimate in any year \((y)\), and \( A_{y+1} \) the estimate in the following year. The value may be positive or negative, depending on the direction of change. The advantage of this metric over some other metrics of assessment variability is that it does not depend on the period of convergence, so showing values on an equal basis over time. It also reflects directly the changes observed by the end-user. The values obtained can be summarized to illustrate the average magnitude of change and the rate of convergence as the root mean square (RMS) of the changes in each subsequent revision. The RMS value is similar in scaling to a standard deviation.

Assessment performance
The results of the historical retrospective analysis for the three main assessment parameters (SSB, \( F_{2-6} \), and \( R \); Figure 2) indicate a major jump in the estimated \( F \) before 1995 and a parallel drop in SSB, but quite consistent estimates thereafter. Downward adjustments in R have been larger in recent years, particularly for 1998, 1999, and 2001. The results of the analytical retrospective are shown as cohort retrospectives in Figure 3. The 1990–1993 cohorts are all overestimated at older ages because of a bias in the selection pattern following the 1995 management changes. Later cohorts are more consistently estimated, tending to converge to a final value as more data become available each year.

The changes in estimates of SSB are perhaps better illustrated in Figure 4, which shows the relative year-to-year revision by subsequent assessments from both the historical and analytical retrospectives. The historical retrospectives (Figure 4a) show that, during the early 1990s, SSB has been revised substantially downwards yearly (and conversely \( F \) has been revised upwards; not shown). In the late 1990s, following the introduction of ICA, annual revisions were more variable from year to year, although they were still substantial. However, since the introduction of inverse variance weighting (2002), the assessment has become much more stable.

Compared with the analytical retrospective, the historical retrospectives show much smaller revisions before 1995 (Figure 4a and b), suggesting that the introduction of ICA with prior weighting and the development of a coherent dataset led to a distinct improvement. The period 1995–2000 also shows more variability in the revisions in the historical retrospective than in the analytical retrospective, although some systematic trends are maintained. During this period, the catches of juvenile herring were dramatically reduced from those of earlier years, and although the attempts to deal with observed changes in the fishery did remove some bias, they also appear to have introduced more noise. The analytical retrospective completely ignores a perceived change in selectivity, and it is apparent that the separable model assumptions used in ICA have difficulties during this period, giving less variability but more bias (subsequent changes are in the same direction). This suggests that current modelling may still be sensitive to the assumptions of a separable catch-at-age model. Therefore, if there are future management changes that result in changes in selection, the assessment may again become biased. In recent years, the variability seems higher in the analytical than in the historical retrospective. Because the two graphs should be almost identical, the obvious explanation is that the reconstruction of the database (particularly the removal of the age-2 index from the acoustic survey) is responsible. Overall, these results suggest that the current assessment methods, the expanding timeseries of data, and possibly more stability in management measures are delivering more stable assessment results, with revisions of close to 5%.

The convergence of the assessments in terms of the RMS value for successive revisions of SSB, averaged over the available data from 1991 to 2007 (Figure 5), shows that, in the historical retrospectives, revisions drop from 16% to a mere 3% within 6 years, whereas revisions in the analytical retrospective drop from 12 to 1%, indicating overall improved convergence.

Analytical retrospective analyses carried out after the removal of one tuning index at a time from the assessment (Figure 6a) indicate that each index contributes to the stability of the assessment, and the use of the combination usually appears to outperform any of the subsets shown. The inclusion of IBTS or MIK does not appear to improve estimates of SSB because neither is particularly informative for this parameter, and their dominant influence is on the recruitment of cohorts (see below). The use of each index on its own (Figure 6b) also gives poorer results than can be achieved by the total set. These comparisons strongly support the conclusion that the weighting of the data is appropriate, because the poorer individual indices have the least impact when removed. However, this conclusion is limited to estimates of SSB and \( F \) (not shown), which had similar properties.

The performance of estimated \( R \) is more complex (Figure 7). Overall, assessment with the full set of indices is the quickest to converge to a final value. However, early convergence is more difficult to evaluate because different indices influence
recruitment in different years. Although excluding the IBTS, or using the MIK on its own, may appear to provide more convergence initially, this is only because of the exclusion of other data on that age class, and models incorporating these options take longer to converge.

Conclusions

A steady improvement in modelling and consistency of data collection, moving to a statistical model in the mid-1990s and incorporating appropriate weights in fitting the data in 2002, has contributed to what now appears to be a consistent and stable assessment. Revisions to SSB and $F$ are currently typically less than 5% in the first year, and decline thereafter. For comparison, the advice based on the assessments has involved changes in stock size and associated TAC advice of between 10 and 50% annually. Therefore, the assessment can provide informative advice that is

Figure 4. Percentage revisions of SSB in subsequent years (first–fifth) by year of assessment using: (a) historical retrospectives, and (b) analytical retrospectives.

Figure 5. Root mean square (RMS, 1991–2006) of the estimates of SSB over the subsequent 10-year period for both historical and analytical retrospectives.
The revisions illustrated here, particularly since 2002, suggest that this may be one of ICES most stable stock assessments.

Acknowledgements

Many people are involved in the current assessment of NS herring. Without all those who carefully and conscientiously collected the data at the markets and on the research vessels, assembled the data and participated in assessments, none of this would have been possible. In particular, I thank Ken Patterson, who first implemented ICA, Dankert Skagen and Martin Pastoors, who provided advice on the development of the current assessment and data weighting, and Niels Daan, Beatriz Roel, and Emma Hatfield for their comments on the draft manuscript.

References


Appendix

Model formulation for North Sea herring assessment

The model used for the 2008 assessment and the analytical retrospectives took the following form:

\[
\begin{align*}
\sum_{a=6, y=1984}^{a=2007, y=2008} C_{a,y} & \cdot \lambda_{\text{MLAI}} \cdot \left( \ln(q_{\text{MLAI}} \cdot \text{SSB}^y) - \ln(\text{MLAI}_y) \right)^2 + \\
\sum_{a=2, y=1992}^{a=2, y=1998} & \lambda_{\text{IBTS}} \cdot \left( \ln(q_{\text{IBTS}} \cdot \tilde{N}_{a,y}) - \ln(\text{IBTS}_{a,y}) \right)^2 + \\
\sum_{a=2, y=1999}^{a=2, y=2007} & \lambda_{\text{ACOUST}} \cdot \left( \ln(q_{\text{ACOUST}} \cdot \tilde{N}_{a,y}) - \ln(\text{ACOUST}_{a,y}) \right)^2 + \\
\sum_{a=1, y=1992}^{a=9, y=2007} & \lambda_{\text{MIK}} \cdot \left( \ln(q_{\text{MIK}} \cdot \tilde{N}_{a,y}) - \ln(\text{MIK}_{a,y}) \right)^2 + \\
\sum_{a=2, y=1990}^{a=2, y=2007} & \lambda_{\text{SSR}} \cdot \left( \ln(\tilde{N}_{a,y}) - \ln(\alpha \text{SSB}_y) \right)^2 .
\end{align*}
\]

Symbols used:

- \(a, y\): age and year
- \(C_{a,y}\): catch
- \(\tilde{N}_{a,y}\): estimated population numbers
- \(\text{SSB}_y\): estimated spawning-stock biomass
- \(\text{ACOUST}_{a,y}\): acoustic index (age-disaggregated \(a = 2–10+)\)
- \(\text{IBTS}_{a,y}\): IBTS index (a = 2–6+)
- \(\alpha\): MIK index (age 1)
- \(q_{a}\): age-dependent catchability
- \(K\): power of catchability model
- \(\lambda_{a}\): age-dependent weighting factor
- \(\lambda_{\text{SSR}}\): stock-recruit relationship weighting factor
- \(\alpha\): age 3 before 1997