Can less be more? Effects of reduced frequency of surveys and stock assessments

Fabian Zimmermann[1,2] and Katja Enberg[1,*]

[1]Institute of Marine Research and Hjort Centre for Marine Ecosystem Dynamics, P.O. Box 1879 Nordnes, N-5817 Bergen, Norway
[2]Department of Biology, University of Bergen, P.O. Box 7803, N-5020 Bergen, Norway

*Corresponding author: tel: +47 41 00 63 68; fax: +47 55 23 85 31; e-mail: katja.enberg@imr.no.

Uncertain and inaccurate estimates are a prevailing problem in stock assessment, despite increasingly sophisticated estimation methods and substantial usage of scientific and financial resources. Annual scientific surveys and assessment group meetings require frequent use of research vessels and skilled research staff and are, therefore, particularly costly. This data- and work-intensive approach is often considered paramount for reliable stock estimates and risk management. However, it remains an open question whether the benefits of increasing assessment effort outweigh its marginal costs, or whether the potential impacts of investing less in assessments could generate net benefits. In this study, we explore how different scenarios of reduced survey and assessment frequencies affect estimated stock biomass, predicted catch, and uncertainty. Data of two Northeast Atlantic stocks, blue whiting (*Micromesistius poutassou*) and Norwegian spring-spawning herring (*Clupea harengus*), and a widely applied stock assessment model are used to compare the impacts of removing surveys and/or annual assessments. The results show that lower survey and/or assessment frequencies tend to result in deviating estimates of spawning-stock biomass and catch and larger confidence intervals, the observed differences being, however, mostly small. While scenarios without a survey datapoint in the assessment year generally produced the largest deviations in estimates, biannual surveys in general did not affect assessment performance substantially. This indicates that a reduced frequency of surveys and assessments could be an acceptable measure to reduce assessment costs and increase the efficiency of fisheries management, particularly when accompanied by thorough management strategy evaluations and risk assessments.

Keywords: fisheries management, scientific surveys, stock assessment, uncertainty.

Introduction

Stock assessment is a fundamental component of fisheries management. Authoritative estimates of current stock biomass and composition are a necessity for predicting future stock development and quota setting. For this purpose, a wide array of stock assessment methods have been developed ranging from simple biomass models that utilize time-series of commercial catch per unit effort to age-structured models that can integrate commercial and scientific data (Hilborn and Walters, 1992; Haddon, 2001). State-of-the-art versions of statistical catch-at-age or cohort analysis models in combination with scientific surveys (for a

review, see Maunder and Punt, 2013) are today considered as the best approximation to true stock sizes and, therefore, dominate stock assessments within the International Council for the Exploration of the Sea (ICES). These methods usually result in estimates of high precision and, potentially, accuracy. However, they are also capital- and work-intensive, particularly when combined with annual surveys. Despite the substantial investments in scientific personnel and equipment, these stock assessment methods also remain plagued by often large uncertainty.

Uncertainty is a general and largely unavoidable issue in stock assessment and advice (Hilborn and Walters, 1992; Patterson

Figure 1. Definition of baseline and scenario assessment runs with modified survey frequency. Configurations for assessment runs of 2014, 2013, and 2008 are shown, but for all years between 2009 and 2012, baseline and scenario assessments were compared following the identical pattern. The configurations were applied to the entire time-series (here just detailed until 2003) of official assessments and surveys (from 1980 and 1988 for blue whiting and NSS herring, respectively, until the endpoint of each scenario, e.g. 2015 for the assessment of 2014) and surveys. In each assessment year (dark grey colour), the entire time-series including a 1-year forecast (light grey) were compared under the assumption that assessments (A) occurred in all previous years and with survey datapoints (S) used in each scenario as detailed. Years without survey in scenarios with reduced survey frequency (Scenarios S+ and S−) are indicated by empty (white) fields.

et al., 2001; Rochet and Rice, 2009; Fromentin *et al.*, 2014). There are two complementary approaches to deal with the problem of uncertainty: (i) the use of assessment models that allow for quantification of uncertainty and its subsequent implementation in management strategy evaluations (Punt, 2015) and risk assessments; and (ii) the attempt to reduce uncertainty with additional information, requiring more and more data, resulting in increasingly complex assessment models. The latter, in particular, leads to data-hungry and work-intensive assessment processes where annual surveys are considered essential. However, the marginal gains in precision from the increasing assessment efforts may be low in view of potential bias in time-series estimates of biomass and fishing mortality and underlying uncertainty because of environmental–ecological processes that cannot be fully understood or predicted. This has led to criticism of the costly assessment process and the necessity of annual surveys, and alternatives such as "cheap and dirty" assessments have been suggested to deal better with data issues and uncertainty (Kelly and Codling, 2006). The required accuracy of stock estimates may also depend on the state of stocks and the harvest strategy (Myrseth *et al.*, 2011). Precautionary harvest control rules (Froese *et al.*, 2011) or multiannual harvest strategies (Marchal and Horwood, 1995; Marchal, 1997) are, therefore, suitable tools to complement reduced assessment effort and compensate for potentially higher uncertainty in stock estimates. This particularly applies for stocks

sizes above precautionary limits and may thus become increasingly relevant with current trends towards improved stock status in well-managed regions (Fernandes and Cook, 2013; Hilborn and Ovando, 2014).

Prior to cost-benefit analysis of reduced assessment and survey frequencies, the impact of such alternative assessment scenarios on stock assessment estimates needs to be quantitatively addressed. To pursue this goal, we explored the deviations from "actual" stock sizes caused by reduced frequency of surveys. We hypothesize that, given the inherent uncertainty of stock assessment, reduced frequency of surveys does not lead to deviations in estimates outside of current confidence intervals (CIs) or to a larger range of CIs. Comparable estimates and CIs would imply that uncertainty may be negligible in the context of the overall high uncertainty, i.e. the assessment uncertainty because of data uncertainty, model assumption, and unaccounted drivers. Furthermore, we investigated the impact of lower assessment frequencies on stock estimates and particularly forecasts of stock size and catch. The present analysis consists, therefore, of three parts: (i) analysing the consequences of removing years of input information, (ii) comparing current assessments with assessments where the survey frequency has been reduced from annual to biannual, and (iii) comparing current assessments with assessments that assume a biannual assessment frequency and thus require a 2-year forecast to cover the period until the next assessment. The

Figure 2. Definition of baseline and scenario assessment runs with modified assessment and survey frequency. Configurations for assessment runs of 2014, 2013, and 2008 are shown, but for all years between 2009 and 2012 baseline and scenario assessments were compared following the identical pattern and for the entire assessed time-series (from 1980 and 1988 to 2014 for blue whiting and NSS herring, respectively). In each assessment year (dark grey colour), the entire time-series including a 2-year forecast (light grey) were compared under the assumption that assessments (A) and surveys (S) occurred in each scenario as detailed. Years without assessment/survey in scenarios with reduced assessment frequency (Scenario A) and reduced assessment and survey frequency (Scenarios AS+ and AS−) are indicated by empty (white) fields. Assessments in years prior to the assessment year are shown (removed) to illustrate the scenario and did not affect the output.

latter considers both a scenario with regular survey frequency and scenarios that combine reduced assessment frequency with reduced survey frequency. Removing or adding years of input information to an assessment can cause methodical changes in assessment estimates that are known as retrospective patterns and typically shift the estimated time-series partly or entirely even if all other input information remains identical. Such patterns are a common phenomenon in stock assessment models and cause estimation differences with every year added or removed (Deroba, 2014; Hurtado-Ferro *et al.*, 2014), a known problem particularly with the assessment of Norwegian spring-spawning herring (*Clupea harengus*) (e.g. see ICES, 2013).

As case examples for evaluating how a shift from annual to biannual assessments and/or surveys influences the historic and forecasted estimates of spawning-stock biomass (SSB) and predicted catch, we used two stocks from the Northeast Atlantic, blue whiting (*Micromesistius poutassou*) and Norwegian spring-spawning (Atlanto-Scandian) herring. Both are large pelagic stocks with complex dynamics as a result of extensive migration patterns in combination with ecological and environmental drivers, notably the abundance and distribution of their prey (Prokopchuk and Sentyabov, 2006; Vikebø et al., 2012; Skagseth *et al.*, 2015), predators (Sætre *et al.*, 2002; Durant *et al.*, 2014; Nøttestad *et al.*, 2015), and competitors (Cabral and Murta, 2002; Langøy *et al.*, 2012) as well as climatic and oceanic conditions (Hátún *et al.*, 2009a, b; Payne *et al.*, 2012). These dynamics lead to fluctuations in stock size that are typically hard to predict and add to the uncertainty in estimates and predictions. The available time-series from the existing stock assessments of both stocks (ICES, 2014) were used as a basis to compare assessments that include complete information with assessments where survey datapoints have been removed. The different assessment scenarios were analysed for the deviation in estimated SSB and predicted catch, which are the basis of management advice and quota decisions.

Methods
Our analyses are based on data and parameter configurations from the ICES stock assessments of blue whiting and Norwegian spring-spawning (NSS) herring (ICES, 2014); for both stocks, we used the stock assessment model SAM (Nielsen and Berg, 2014). SAM is currently the model used for the ICES assessment of the blue whiting stock, whereas the NSS herring stock is assessed with
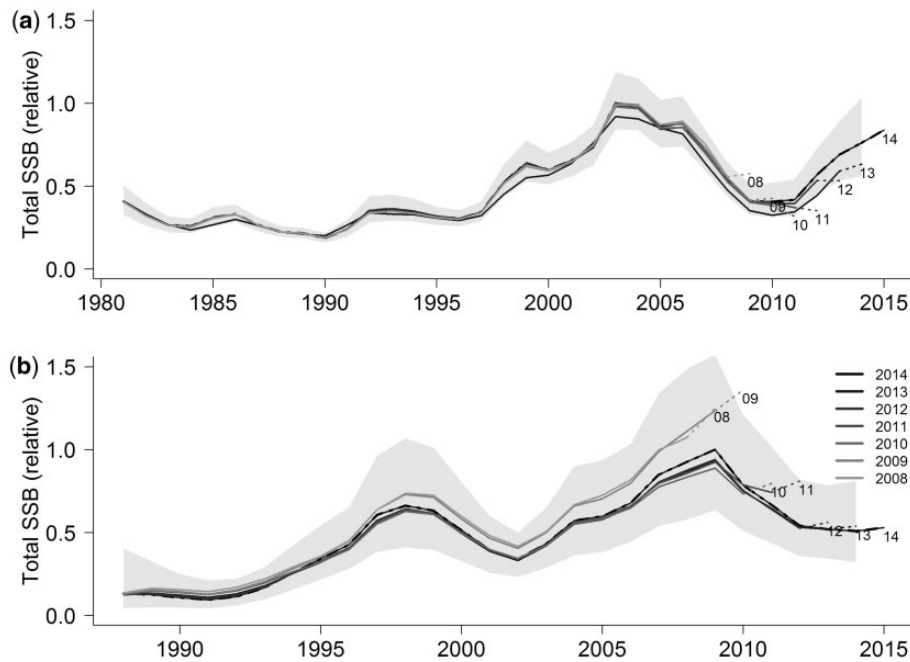
Figure 3. Time-series of spawning-stock biomass (SSB) for baseline runs of 2008–2014 (greyscale) over the whole assessment period for blue whiting (a) and Norwegian spring-spawning herring (b). Solid lines denote time-series of assessment estimates, dotted lines indicate forecasted periods. All values were standardized to maximum total SSB of the latest (2014) assessment, and light grey areas indicate the 95% confidence intervals of 2014.
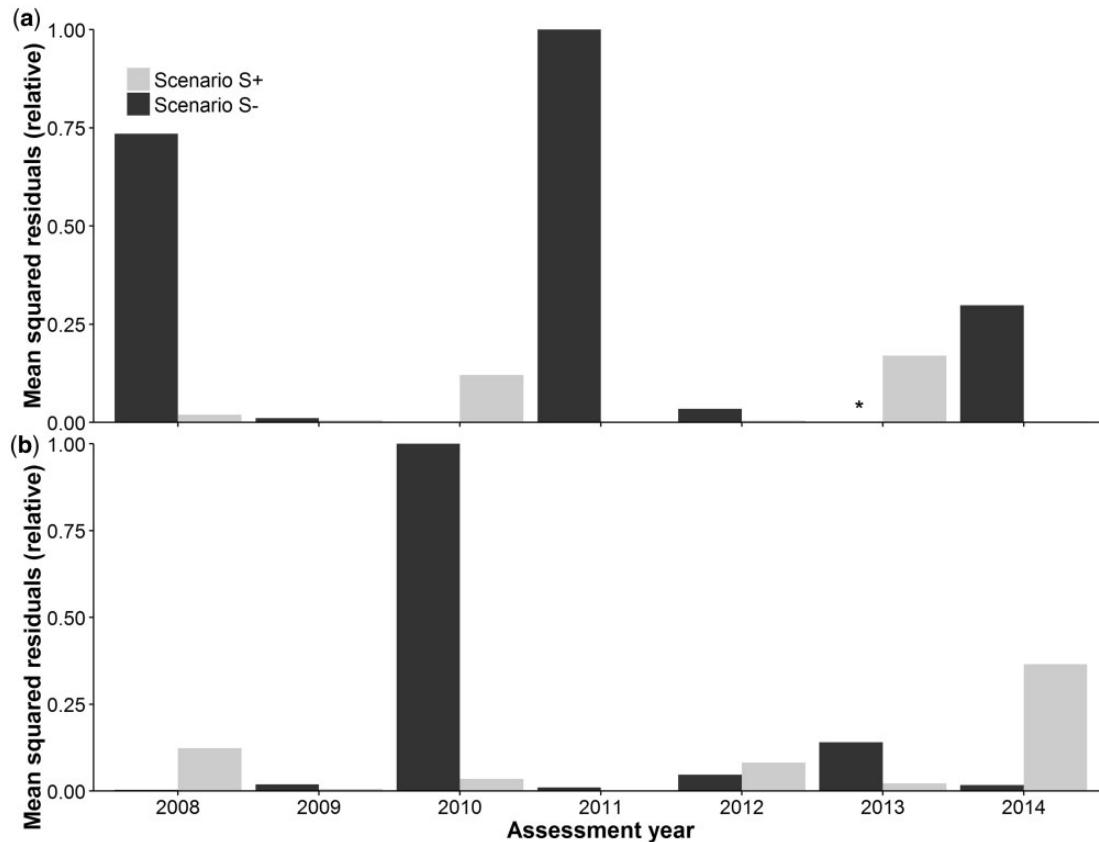
Figure 4. Relative mean squared residuals of spawning-stock biomass between baseline and survey scenario runs of all assessments from 2007 to 2014 for blue whiting (a) and Norwegian spring-spawning herring (b). In scenario runs, the assessments were made with the frequency of surveys reduced to once every other year, either with a survey in the assessment year (Scenario S+) or without (Scenario S−). Squared residuals were calculated from the deviation between SSB in baseline and scenario in each year of the assessed time-series and averaged over the whole time-series. All values were standardized to their stock-specific maximum. Scenario S− of blue whiting in 2013 (*) was removed.
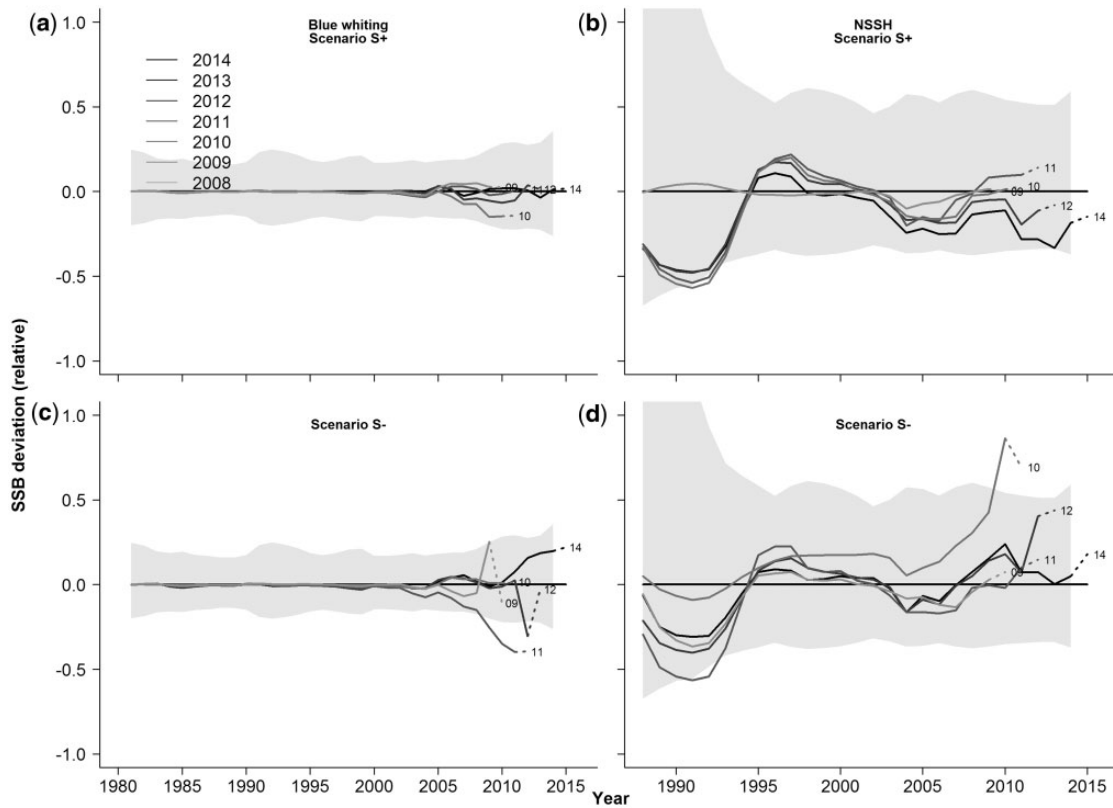
Figure 5. Relative deviation between spawning-stock biomass (SSB) of survey scenario runs from 2008 to 2014 and corresponding baseline runs for blue whiting (a and c) and Norwegian spring-spawning herring (b and d). In scenario runs, the assessments were made with the frequency of surveys reduced to once every other year, either with a survey in the assessment year (Scenario S+) or without (Scenario S−). Deviation is calculated as the proportional difference between annual SSB of scenario and baseline runs with 0 = no deviation, >0 = higher SSB in scenario run, and <0 = lower SSB in scenario run. Grey areas indicate the deviation between the point estimate of SSB and upper and lower boundaries of the 95% confidence intervals in the baseline 2014 assessment.

a VPA model TASACS (ICES, 2014). To conduct the analysis within the same model framework, the assessment of NSS herring was here carried out using SAM and applying the data and parameter configuration as in the official assessment reports (ICES, 2014), however with modified survey information. The official assessment of NSS herring incorporates seven current or historic surveys with time-series of varying length, target ages, and totality that all provide information and/or tuning indices. To avoid a multidimensional and therefore complex analysis, we excluded all surveys except the international acoustic surveys on the feeding areas in the Norwegian Sea in May (IESNS; ages 3–15+ years). The IESNS provides the most complete and consistent overview of the adult portion of the stock from 1996 until 2014 and is the backbone of the NSS herring assessment. All further modifications of survey indices were made as defined by the scenario runs (Figures 1 and 2). Besides survey indices, input information consisted of time-series of catch at age, mean weights at age from catch and stock, maturity ogives, and natural mortality, as detailed in the official assessment report (ICES, 2014). This input information is partially based on survey data, but typically averaged over several years and, therefore, not affected by changes in survey frequency.

Our analysis used the assessment framework described above as baseline and was conducted in three steps: (i) comparison among baseline runs to reveal retrospective patterns, (ii) comparison of baseline runs to scenario runs that assumed a reduced frequency of surveys, and (iii) comparison of baseline runs to scenario runs that assumed reduced frequency of assessments or a combination of reduced frequency of assessments and surveys. Retrospective patterns were analysed by systematically removing years of input information between the years 2014 and 2008, thus revealing deviations and potential biases among the entire time-series of baseline runs. In baseline runs, assessment and surveys were carried out annually, corresponding to the current default for both of these stocks. The scenario runs were divided into two groups: one that kept the assessment frequency intact at an annual level, but focused on reduced survey frequency, and one that evaluated the impacts of a reduction in assessment frequency. In the first group (Figure 1), we assumed that surveys were reduced to bi-annual frequency, with the last survey occurring either in the assessment year *a* (Scenarios S+) or in the year *a*−1 before the assessment (Scenario S−), and analysed how estimates deviated between each scenario and the baseline. In the second group (Figure 2), we assumed a reduced assessment frequency to study the effects on forecasting periods.

Short-term forecasts are an essential part of management advice that covers the period until the next assessment and provides the basis for catch advice and quota regulations. While 1-year forecasts are sufficient to cover the whole period until the next assessment for annual assessments, biannual assessments require
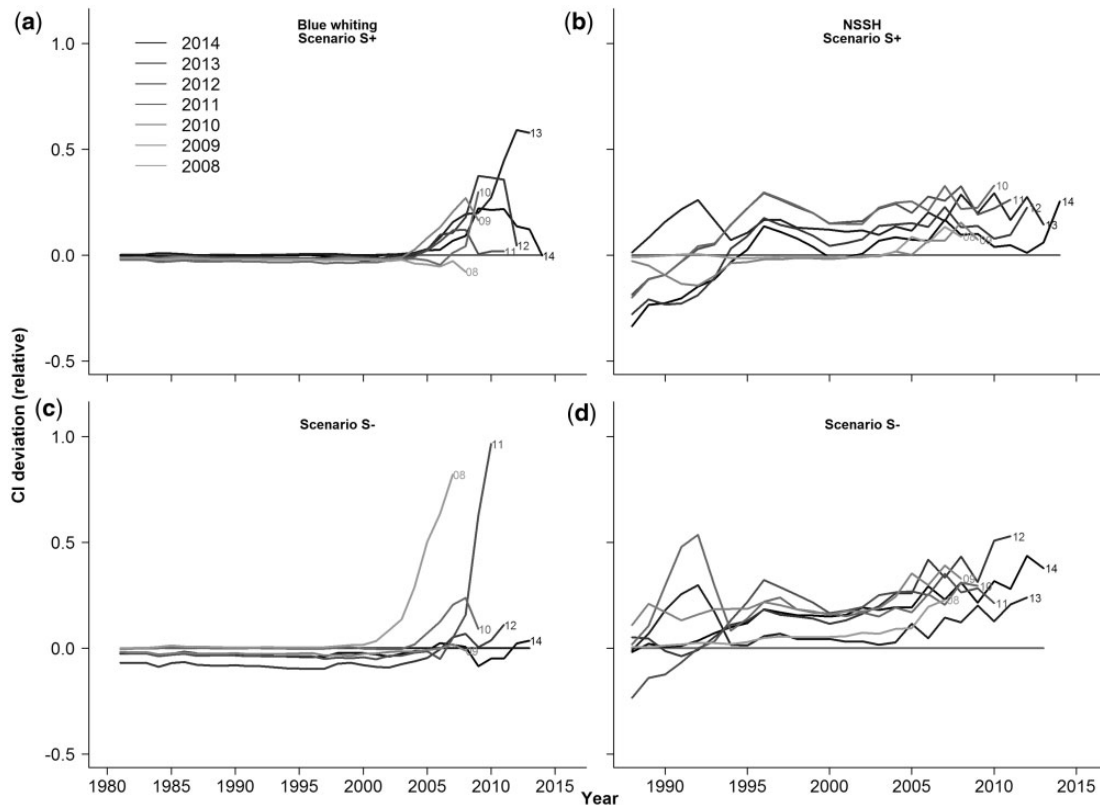
Figure 6. Annual deviation between relative range of confidence intervals (CI) of spawning-stock biomass (SSB) from survey scenario and baseline runs for all runs from 2008 to 2014 (greyscale) in each scenario for blue whiting (a and c) and Norwegian spring-spawning herring (b and d). In scenario runs, the assessments were made with the frequency of surveys reduced to once every other year, either with a survey in the assessment year (Scenario S+) or without (Scenario S−). The range between lower and upper 95% CIs was standardized to the corresponding SSB estimate of each year and run. Deviation is calculated as the difference between annual standardized CIs of scenario and baseline runs with 0 = no deviation, >0 = larger CI range in scenario run, and <0 = smaller CI range in scenario run.

forecasting the stock development 2 years into the future. That is, if assessments only occur biannually, forecasted stock estimates are necessary from the assessment year a into the following 2 years, i.e. $a+1$ and $a+2$, to provide advice on catches until the next assessment in $a+2$, in contrast to 1-year forecasts in regular (baseline) assessment. Because in baseline runs, assessments are conducted annually, forecasts for year $a+2$ can be made in the assessment in $a+1$, whereas in scenario runs, the assessment in year a needs to provide forecasts until $a+2$. Accounting for this length difference in forecast periods, we analysed the stock assessment time-series from the terminal forecast year $a+2$ (when the next assessment would occur) backwards, comparing scenario runs in a with baseline runs in the following year $a+1$ according to their specific length of forecast periods of 2 years and 1 year, respectively. The second group contains one scenario where only assessment frequency was reduced, but survey frequency remained as in the baseline (Scenario A), and two scenarios that combined biannual assessments with biannual surveys where the last survey occurred either in the assessment year t (Scenarios AS+) or in the year $a-1$ before the assessment (Scenario AS−).

The geometric mean of historic recruitment has been used as input recruitment in all forecasts. This ignores that, in reality, adjusted recruitment forecasts are used for blue whiting if sufficient observations from other surveys allow for a prediction of higher-than-average recruitment in the corresponding year

(ICES, 2014). In all forecasts, we used the fishing mortality of each specific year that was assessed in the final baseline assessment of 2014. The only exceptions were forecasts beyond 2014, in which the fishing mortality corresponding to each stock's management plan ($F = 0.18$ for blue whiting and $F = 0.125$ for NSS herring) was selected as input fishing mortality in all baseline and scenario runs. For simplification, predicted catches were assumed to fully correspond to total catch quotas and realized catches, implying complete compliance of regulators and fishers with advised catches.

In both scenario groups, we quantified the deviation between scenario and baseline runs from the resulting outputs of SSB and predicted total catches over the entire assessed time-series and forecasted years, respectively. For survey scenarios (Scenarios S+ and S−), deviations were calculated as proportional deviation between SSB in each scenario S and baseline B in all years t:

$$D_{S,a}^{SSB}(t) = \frac{SSB_{S,a}(t)}{SSB_{B,a}(t)} - 1 \tag{1}$$

where a denotes the year when each assessment run is performed, whereas years t represent all years within the assessed time-series of length T, i.e. $t = 0, 1, \ldots, T, T+1$ (with year T corresponding to year a and $T+1$ the forecasted estimates for the year following assessment). For both stocks, the years 2008 until 2014 were
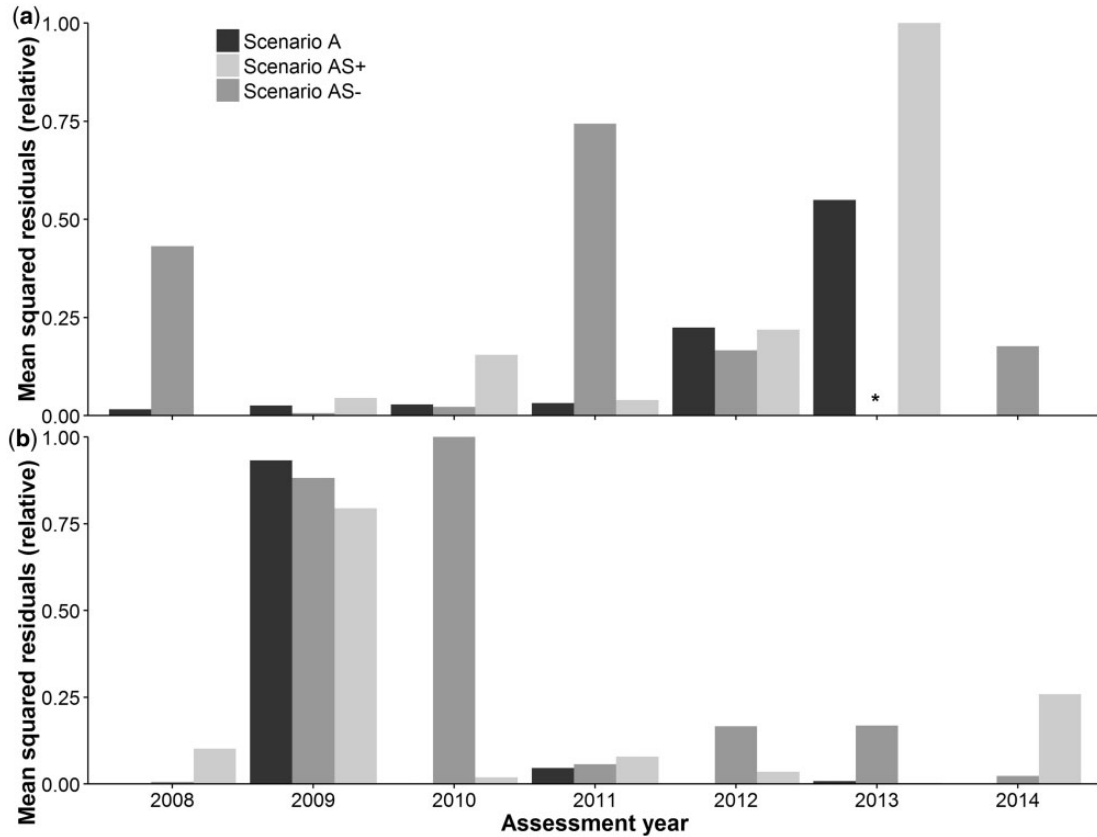
Figure 7. Relative mean squared residuals of spawning-stock biomass between baseline and assessment scenario runs of all assessment years from 2008 to 2014 for blue whiting (a) and Norwegian spring-spawning herring (b). In scenario runs, the frequency of assessments was reduced to once every other year, either with annual surveys (Scenario A) or biannual surveys that coincide with the assessment year (Scenario AS+) or not (Scenario AS−). Residuals were calculated as the deviation between SSB in baseline and scenario in each year of the assessed time-series. All values were standardized to their stock-specific maximum. Scenario S− of blue whiting in 2013 (*) was removed.

used as assessment years a. To quantify total deviation between each scenario and baseline run independent of their direction (i.e. whether deviations are positive or negative), we used mean squared residuals:

$$MSR_{S,a}^{SSB} = \frac{\sum_{i=1}^{T+1} (SSB_{B.a}(t) - SSB_{S.a}(t)\ )^2}{T+1} \quad (2)$$

for each assessed time-series of $T+1$ years. Relative confidence intervals CI were calculated as the range between the upper and lower 95% CIs of SSB standardized to the point estimate of SSB:

$$CI_a(t) = \frac{SSB_a^{upper}(t) - SSB_a^{lower}(t)}{SSB_a(t)}, \quad (4)$$

and compared between scenario and baseline runs:

$$D_{S,a}^{CI}(t) = \frac{CI_{S,a}(t)}{CI_{B,a}(t)} - 1. \quad (5)$$

The same analysis was repeated for the assessment scenarios (Scenarios A, AS+, and AS−) in the second group. The only difference here was to adjust how the assessment years were compared with account for the required 2-year forecast period under biannual assessment frequency. Accordingly, deviations and

residuals were calculated between each assessment in year a of all scenario runs and the corresponding baseline runs of assessment in year $a+1$. Specifically, this leads to the following equations:

$$D_{S,a}^{SSB}(t) = \frac{SSB_{S,a}(t)}{SSB_{B,a+1}(t)} - 1 \quad (6)$$

and

$$MSR_{S,a}^{SSB} = \frac{\sum_{i=1}^{T+2} (SSB_{B.a}(t) - SSB_{S.a+1}(t)\ )^2}{T+2} \quad (7)$$

From forecasted catches C in all years $a+1$ of each scenario, mean absolute and mean squared residuals over the total number of assessment runs from 2008 until 2014, i.e. 7, in each scenario were calculated:

$$MAR_S^C = \sum_1^N \frac{C_{B,a}(T+1) - C_{S,a}(T+1)}{7} \quad (8)$$

and

$$MSR_S^C = \sum_1^N \frac{(C_{B,a}(T+1) - C_{S.a}(T+1)\ )^2}{7}, \quad (9)$$
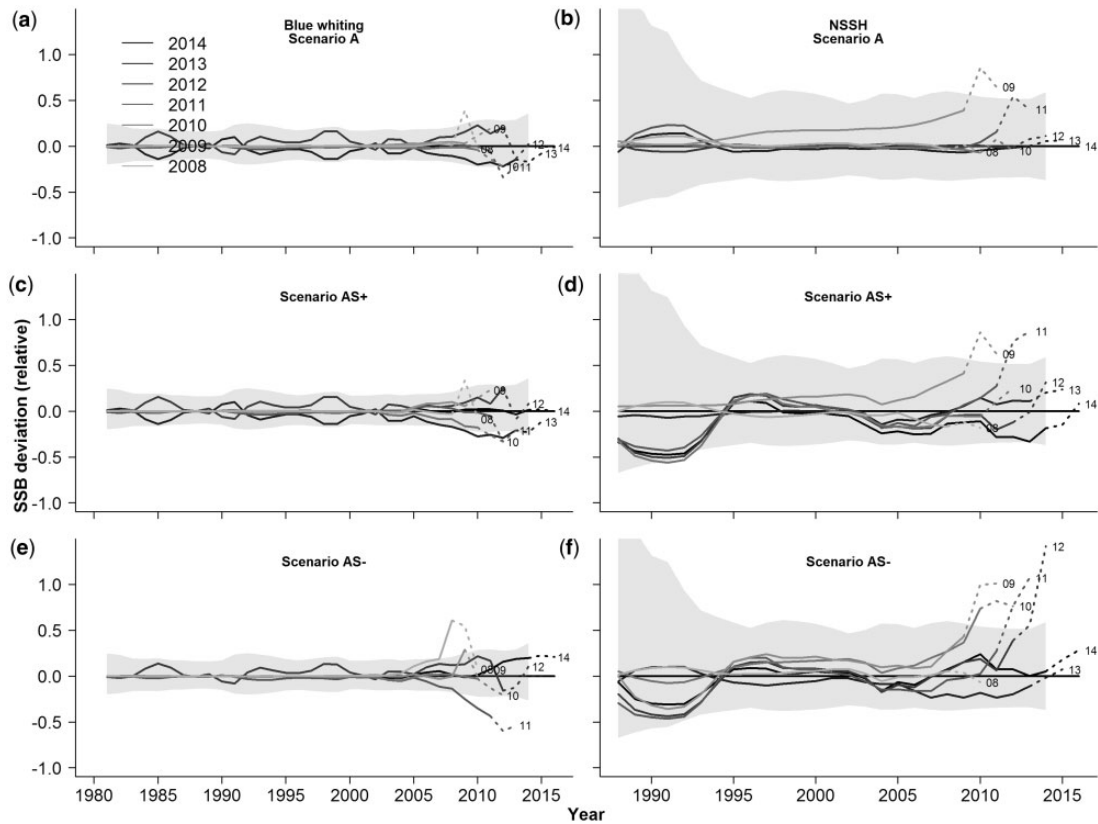
Figure 8. Relative deviation between spawning-stock biomass (SSB) of assessment scenario runs from 2008 to 2014 (greyscale) and corresponding baseline runs for blue whiting (a, c, and e) and Norwegian spring-spawning herring (b, d, and f). Solid lines denote time-series of assessment estimates, dotted lines indicate forecasted periods. In scenario runs, the frequency of assessments was reduced to once every other year, either with annual surveys (Scenario A) or biannual surveys that coincide with the assessment year (Scenario AS+) or not (Scenario AS). Deviation is calculated as the proportional difference between annual SSB of scenario and baseline runs with 0 = no deviation, >0 = higher SSB in scenario run, and <0 = lower SSB in scenario run. Grey areas indicate the deviation between mean SSB and upper and lower boundaries of the 95% confidence intervals in the baseline 2014 assessment.

with $C_B = C_a$ and $C_S = C_a$ for scenarios with reduced frequency of surveys (Scenarios S+ and S−) or $C_B = C_{a+1}$ and $C_S = C_a$ for scenarios with reduced frequency of assessments (Scenarios A, AS+, and AS−). This means that, from each scenario assessment in year a, the catch C_S was predicted 1 year ahead into advice year $t = T + 1$ (the year following a) and compared with the predicted catch C_B in the same year t from the baseline assessment in $a + 1$. To exemplify, given a scenario assessment in 2013, catch C_S for 2014 was calculated and compared with C_B of 2014 as predicted in the 2014 baseline assessment.

Results

Retrospective analysis of the baseline assessments

We first carried out a retrospective analysis by comparing the baseline runs from different assessment years, revealing retrospective patterns and variation in SSB among the baseline runs (Figure 3). No clear trend could be detected in blue whiting (Figure 3a). From all baseline assessments, 2013 showed the largest deviation in SSB for blue whiting compared with the latest (2014) assessment run, albeit it remained within the 95% CIs of the 2014 run. On the other hand, retrospective patterns were found in NSSH with increasing deviation in SSB compared with the latest (2014) assessment the farther in the

past an assessment was set, and particularly 2008 and 2009 stand out (Figure 3b). Beyond 2008 and 2009, the deviations were only minor and consisted mostly of misdirected forecasting trends in specific years compared with assessment run using the full time-series of data up to 2014. In both stocks, forecasted years in several assessments failed to capture the actual observed trend.

Reduced survey frequency

For reduced survey frequency, the deviation in SSB between baseline and scenario runs varied clearly between assessment years (Figure 4); the variation in SSB remained, however, mostly within the lower and upper boundaries of the 95% CIs of the 2014 baseline run (Figure 5). Analysis of mean squared residuals (MSR) between baseline and scenario runs showed that a few specific combinations of scenario and assessment years resulted in major deviations compared with all others (Figure 4). For blue whiting, the assessment of 2013 that used biannual survey frequency without a survey in the assessment year (Scenario S−) was removed because of model converging problems for this particular assessment and scenario, with subsequently strongly diverging SSB values and large increases in uncertainty. In NSSH, biannual survey frequency without a survey in the
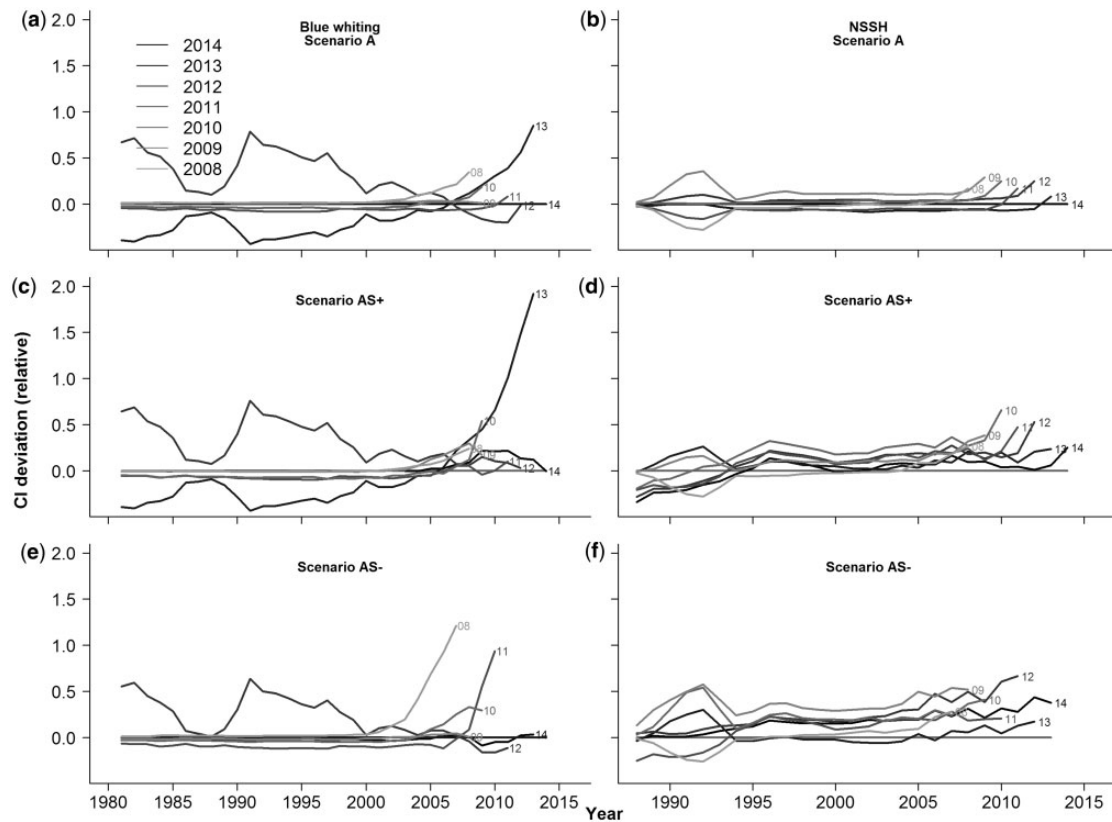
Figure 9. Annual deviation between relative range of confidence intervals (CI) of spawning-stock biomass (SSB) from assessment scenario and baseline runs for all runs from 2008 to 2014 in each scenario for blue whiting (a, c, and e) and Norwegian spring-spawning herring (b, d, and f). In scenario runs, the frequency of assessments was reduced to once every other year, either with annual surveys (Scenario A) or biannual surveys that coincide with the assessment year (Scenario AS+) or not (Scenario AS−). The range between lower and upper 95% CIs was standardized to the corresponding SSB estimate of each year and run. Deviation is calculated as the difference between annual standardized CIs of scenario and baseline runs with 0 = no deviation, >0 = higher CI in scenario run, and <0 = lower CI in scenario run.

assessment year of 2010 resulted in the largest mean squared residuals compared with all other runs, underlining a general tendency for higher MSR in the scenario with biannual survey frequency without a survey in the assessment year (Scenario S−) compared with those with a survey in the assessment year (Scenario S+). There are, however, exceptions of higher MSR in runs with biannual survey frequency with a survey in the assessment year than in those without, notably in the assessments of 2008, 2012, and 2014 in NSSH.

The spawning-stock biomass in all scenario runs of biannual survey frequency remained mostly within the 95% CIs of the baseline assessment (Figure 5), however, with a trend towards higher uncertainty reflected by proportionally larger CIs around the SSB when survey frequency is reduced (Figure 6). In both stocks, deviations outside of the baseline CIs could be observed in the forecasted final years of each assessment, where an increase in uncertainty was to be expected. The proportional range of CIs tended to be larger in most scenario runs compared with their baseline, with few exceptions, particularly for early years of the NSSH time-series. Overall, scenarios with biannual survey frequency tend to result in larger increases in CI range. While in blue whiting specific assessment runs stand out with strong increases in uncertainty for the final years of their assessment period, most assessment runs follow a similar pattern in NSSH.

Reduced assessment frequency

Reducing the frequency of assessments resulted in very minor deviations in most cases (Figure 7). When assessment frequency is reduced, but everything else is left identical to the baseline (Scenario A), deviations could be observed mostly for the years 2012 and 2013 in blue whiting and 2009 in NSSH. The 2013 assessment runs in blue whiting generally show the largest deviations and fail entirely for the scenario with biannual survey frequency (Scenarios S− and AS−), whereas for NSSH, year-specific deviations occurred in all scenarios in 2009 as a consequence of the retrospective patterns in baseline assessments of 2008 and 2009 (Figure 3). In general, when biannual assessments were applied to scenarios with reduced survey frequency, the resulting patterns in MSR reflected the ones in the scenarios with reduced survey frequency and annual assessments (Scenarios S+ and S−) (Figure 4). This suggests that deviations in scenarios with reduced assessment and survey frequency (Scenarios AS+ and AS−) were largely driven by biannual surveys and only to a lesser extent by biannual assessments. There are, however, several exceptions in NSSH where lower assessment frequency increased MSR of specific assessment years in scenarios with reduced survey and assessment frequency (AS+ and AS−) compared with those with reduced survey frequency alone (Scenarios S+ and S−). Within assessment runs, the strongest deviations occurred for the final years, particularly
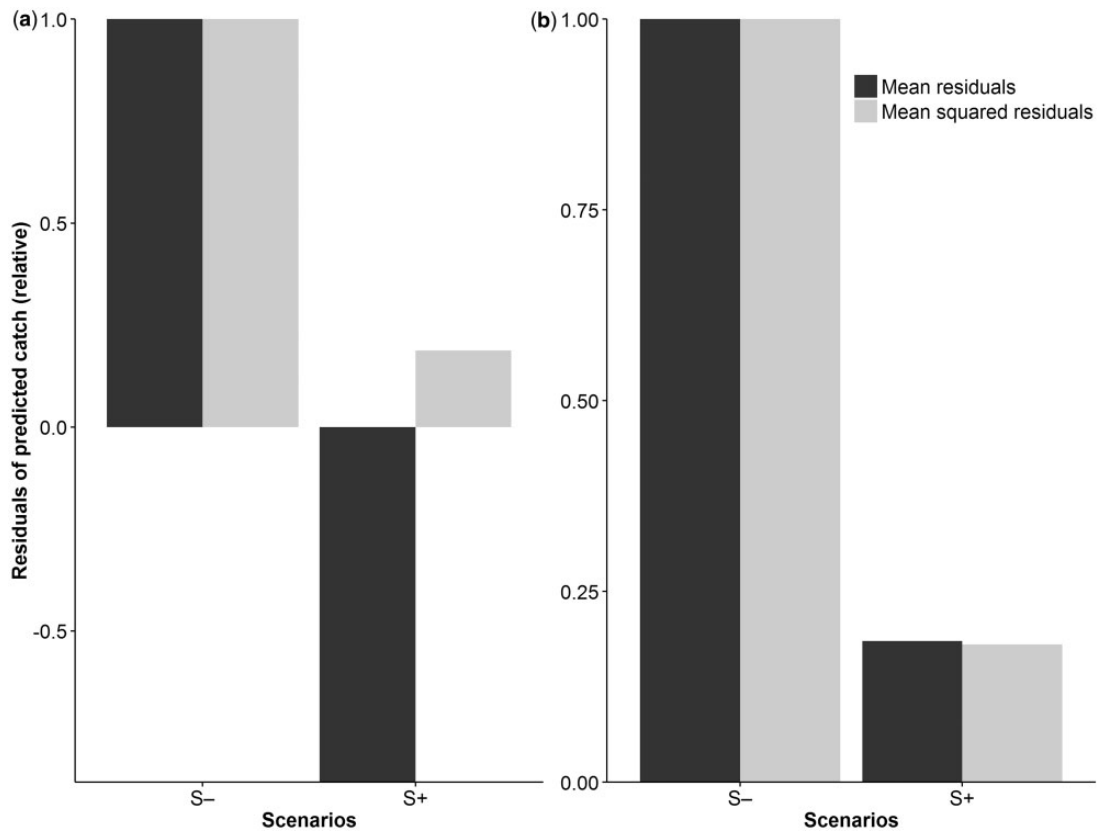
Figure 10. Mean squared residuals (ignoring the direction of the deviations) and mean residuals (taking into account the direction of the deviations) of predicted total catch in the year following the assessment year between survey scenario and baseline runs for blue whiting (a) and NSSH (b), averaged overall assessment years. Positive (negative) values indicate higher (lower) catch predictions in the scenario runs compared with the baseline run with annual survey and assessment. All values were standardized to their stock-specific maximum.

the forecasted years (Figure 8). In both stocks and all three scenarios, this led to estimates outside the 95% CIs of the baseline assessment from 2014 and an increase in uncertainty (Figure 9). Again, this reflected the patterns observed for scenarios with reduced survey frequency and annual assessment (Scenarios S+ and S−); however, the biannual assessments resulted in additional deviation of SSB estimates and their CIs, as illustrated by the scenario with reduced assessment frequency and annual surveys (Scenario A).

Implications of reduced survey and assessment frequency for catch

Residual analysis of predicted total catches between scenario and baseline runs revealed a uniform trend in the total difference for both stocks, but clear stock-specific patterns in the direction of the differences (Figures 10 and 11). MSR were used as an indicator for the total difference in the predicted catches and underlined the results on deviations in SSB. Differences in total catches were mostly driven by reduced survey frequency without a survey in the assessment year (Scenario S−) and amplified further by reduced assessment frequency. Overall, the highest MSR in both stocks were, therefore, found for the scenario with biannual assessment and survey frequency and without a survey in the assessment year (Scenario AS−, Figure 11). Catch predictions confirmed the general trend towards highest deviations and

uncertainty in those two scenarios without a survey in the assessment year.

For blue whiting, deviations in predicted catches were less pronounced and clear trends among the assessment years were lacking (Figure 11a). Residuals of predicted catches overall assessment years were higher in scenario runs without a survey datapoint in the assessment year (Scenarios S− and AS−) or with only reduced frequency of assessments (Scenario A) compared with the baseline, but lower for biannual assessment and surveys taking place in the same year (Scenario AS, Figure 11a). For NSSH on the other hand, predicted catches were higher in most of the assessment years than in the baseline (Figures 10b and 11b). Accordingly, mean residuals of predicted catches that include information on the direction of deviations, i.e. whether deviations are positive or negative, were in all cases positive and reflected the identical pattern as mean squared residuals. This means that the predicted catches for NSSH were overestimated in all scenarios with reduced assessment frequency compared with the baseline. The multiannual forecast required with reduced assessment frequency was giving too positive a view of future catches. This applies in particular for scenarios S− (Figure 10) and AS− (Figure 11) which lack a survey datapoint in the assessment year. The deviations in predicted catches were a direct result of trends in SSB estimates and, therefore, reflected the tendency to under- or overestimate predicted SSB in blue whiting and NSSH, respectively.
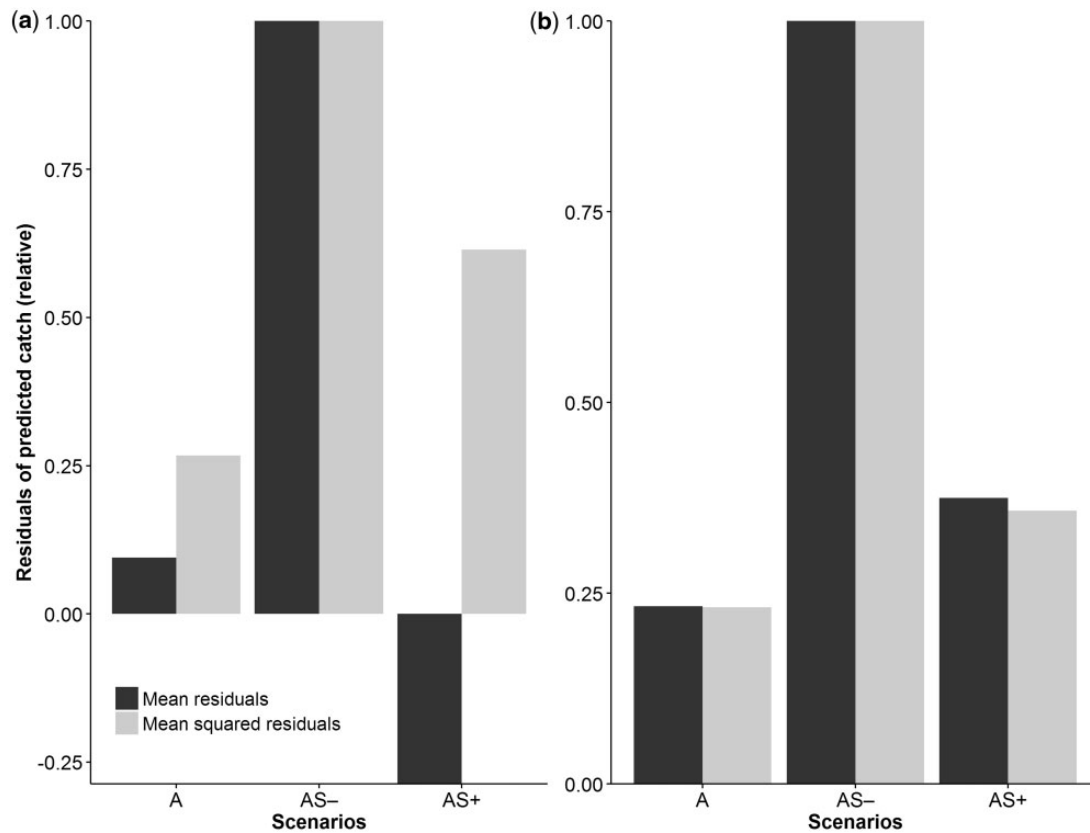
Figure 11. Mean squared residuals (grey; ignoring the direction of the deviations) and mean (absolute) residuals (black; taking into account the direction of the deviations) of predicted total catch in the year following the assessment year between assessment scenario and baseline runs for blue whiting (a) and NSSH (b), averaged overall assessment years. Positive (negative) values indicate higher (lower) catch predictions in the scenario runs compared with the baseline run with annual survey and assessment. Scenarios of assessment year a were compared with baseline assessments of assessment year $a + 1$. All values were standardized to their stock-specific maximum.

Discussion

Our quantitative analysis of different survey and assessment frequencies shows, within a realistic stock assessment framework, that lower frequencies may be an option to reduce overall costs associated with estimating stock size. Our case studies of blue whiting and NSS herring illustrate that reducing survey and/or assessment frequencies results in small deviations in assessment estimates that lay within the CIs for most scenarios and assessment years. The specific performance of different scenarios is stock- and year-specific; however, reducing survey frequency generally had a larger impact on the deviation of estimates than reduced assessment frequency, and scenarios without surveys in the assessment years showed the clearest deviations. This indicates that, with biannual surveys, the years without datapoints are, quite expectedly, the most problematic. This should be taken into account when considering overall management strategies, and its consequences need to be tested with management strategy evaluations. Nevertheless, the results partially confirmed our hypothesis that survey and assessment frequency may be reduced without estimates deviating outside of CIs or increases in uncertainty. Biannual surveys and assessments in the same year did not necessarily deviate more than biannual (annual) assessments combined with annual (biannual) surveys and were close to baseline estimates, although with a tendency towards increasing

deviation in forecasted estimates. This suggests that these scenarios should be considered with necessary caution in the future as alternative assessment strategies.

Within the trade-off between high survey and assessment costs and the need for accurate estimates, moderately and adequately reduced survey and assessment frequencies could provide a good solution for some stocks. As previously concluded (ICES, 2012), this applies specifically to stocks with low harvest pressure and robust stock assessments, and/or which are in a good state (i.e. above MSY biomass or precautionary biomass reference points) and with stable population dynamics (i.e. not very short-lived species with a population structure dominated by recent recruits). Reducing the assessment expenses could also provide an additional incentive for a further spread of robust harvest control rules and multiannual management strategies (Marchal, 1997; Kell *et al.*, 2006), aiming at stable catches and stock sizes above precautionary limits to mitigate potentially higher risks of overfishing under biannual assessments. Our results show that the precise effects of lower assessment or survey frequency depend on the stock and available data, underlining that alternative assessment scenarios should be evaluated in stock-specific management strategy evaluations (ICES, 2012). Furthermore, bioeconomical analysis will be necessary to explore whether reduced survey and assessment frequency could create benefits that outweigh the possibility of higher uncertainty in specific stocks. Our results

underline that catch predictions are ultimately driven by the potential bias in SSB estimates, which means that they can under- or overestimate the recommendable catches for management advice. Both situations have potential economic repercussions, either in the form of overfishing that requires compensation with lower catches in the future or through underfishing that underutilizes stock productivity. To determine the economic consequences in detail requires extensive simulations. Based on the present results, further research and case studies seem recommendable, particularly for well-managed stocks above precautionary limits as well as stocks with generally low catches or of little economic value that put an investment into annual assessments into question.

Large uncertainty is a common problem in stock assessment, and recruitment estimates are typically a major source of uncertainty despite extensive survey data. For stocks with high recruitment variability because of little understood physical conditions, such as NSS herring (Skagseth *et al.*, 2015), unusually weak or strong year classes are very difficult to predict, but can determine stock size for years. Recruitment uncertainty is, therefore, a major driver of overall uncertainty and may explain the large CIs for NSS herring in particular, which are between 50% and 100% of estimated stock size for most parts of the time-series, as well as the tendency to miss most recent trends that were revealed in our retrospective analysis. As a consequence of this underlying uncertainty, removing survey datapoints results in minor deviations in stock estimates within the 95% CIs of the baseline assessment and relative changes of CIs between −50% and 50% in the conservative scenarios with annual assessments (Scenario S+) or annual surveys (Scenario A). A similar effect occurs when both survey and assessment are undertaken biannually (Scenario AS+). This holds true as long as robust and sufficient datapoints from surveys are available for the assessment year and within the time-period closest to the assessment year. If these conditions are not satisfied, stock estimates and CI tend to deviate more, illustrated by scenarios without a survey in the assessment year (S− and AS−), and particularly for blue whiting, where the 2010 survey indices are not included in the blue whiting assessment, which leads to several consecutive years without survey information for various combinations of scenarios and assessment years.

With lower frequencies of surveys or assessments, the relative weight of a single year of survey data and assessment increases. As a consequence, the fewer survey datapoints are available or the less assessment-based advice is provided, the more important the reliability and precision of existing ones becomes. A reduction in survey or assessment frequency also tends to decrease the acceptable margin of error and, therefore, increase the associated risk of reducing a stock below precautionary limits. If a survey fails or produces questionable indices, for instance, the relative impact is more dramatic within a biannual survey routine compared with an annual one. Our study does not incorporate such scenarios and the associated uncertainty effects, but the impact would generally be increased uncertainty in the assessment and consequent reduction in acceptable catch levels. The same issue is of particular relevance in predictions of SSB and catch, which are the basis for management advice, but tend to be incapable of capturing current trends in stock development in periods of increasing or decreasing SSB. This problem is reflected in the stock-specific over- and underestimation of predicted catches and may be emphasized by biannual assessments that depend on forecast for 2 years and skip a year of possible adjustments to trend estimates.

However, an adaptive survey and assessment strategy that increases the frequency when required may mitigate this problem.

The conclusions of our study are limited by the use of a specific stock assessment model with its assumptions and that we remove years in hindsight without accounting for possible changes in management decision because of deviating estimates. Assessment models, as all models, are only an approximation to reality and are constrained by the quality of data to which they are fitted. This is particularly relevant for the baseline assessments in our study, which correspond to the actual assessments and were thus treated as "true" values. However, those estimates are not necessarily more accurate than those from the scenario runs and should not be confused with data (Brooks and Deroba, 2015). Accordingly, deviations from baseline runs do not necessarily indicate deviations from reality. As the latter is not known and the estimated stock sizes in fisheries management are treated as sufficiently "true" values, they subsequently become the most plausible benchmark for the performance of scenario runs.

Analysing assessment data in hindsight, i.e. retrospectively, may cause potential biases through assumptions on fishing mortalities and the way baseline and scenario runs are compared. As biannual assessments require forecasts of 2 years to cover the period until the next assessment, we aligned scenarios with biannual assessments A, AS+, and AS− to baseline runs from the following assessment year. If retrospective patterns (Deroba, 2014; Hurtado-Ferro *et al.*, 2014) caused deviations between assessment years, these deviations became compounded with those caused by the scenario modifications. Consequently, such effects may lead to deviations in scenario runs that are not actually caused by the assessment or survey frequency. Furthermore, we assumed that predicted catches are based on the same fishing mortalities as they occurred according to the final assessment of 2014 (ICES, 2014). This ignores that managers could set different TACs based on the scenario assessment and limits, therefore, their potential impacts, particularly for scenarios where a 2-year forecast would be necessary. To fully account for this, either a strict and well-defined harvest control rule or a broad evaluation of all potential TAC decisions would be required. The former did not exist for both stocks over the whole period analysed (although such a harvest control rule is in place in today's management of NSS herring), the latter would substantially increase the complexity of the analysis. Accordingly, we chose to study the key question in a simple, comparable, and fundamental way over a representative range of different scenarios and assessment years. As a next step, we suggest combining alternative assessment scenarios with management strategy evaluations in specific case studies to investigate the interplay of frequency of surveys and assessments, assessment uncertainty, and management decisions.

Acknowledgements

References

Brooks, E. N., and Deroba, J. J. 2015. When "data" are not data: the pitfalls of *post hoc* analyses that use stock assessment model output. Canadian Journal of Fisheries and Aquatic Sciences, 72: 634–641.

Cabral, H., and Murta, A. 2002. The diet of blue whiting, hake, horse mackerel and mackerel off Portugal. Journal of Applied Ichthyology, 18: 14–23.

Deroba, J. J. 2014. Evaluating the consequences of adjusting fish stock assessment estimates of biomass for retrospective patterns using Mohn's Rho. North American Journal of Fisheries Management, 34: 380–390.

Durant, J. M., Skern-Mauritzen, M., Krasnov, Y. V., Nikolaeva, N. G., Lindstrøm, U., and Dolgov, A. 2014. Temporal dynamics of top predators interactions in the Barents Sea. PLoS One, 9: e110933.

Fernandes, P. G., and Cook, R. M. 2013. Reversal of fish stock decline in the Northeast Atlantic. Current Biology, 23: 1432–1437.

Froese, R., Branch, T. A., Proelß, A., Quaas, M., Sainsbury, K., and Zimmermann, C. 2011. Generic harvest control rules for European fisheries. Fish and Fisheries, 12: 340–351.

Fromentin, J.-M., Bonhommeau, S., Arrizabalaga, H., and Kell, L. T. 2014. The spectre of uncertainty in management of exploited fish stocks: the illustrative case of Atlantic bluefin tuna. Marine Policy, 47: 8–14.

Haddon, M. 2001. Modelling and Quantitative Methods in Fisheries. CRC Press, Boca Raton, FL. 424 pp.

Hátún, H., Payne, M., Beaugrand, G., Reid, P., Sandø, A., Drange, H., Hansen, B., *et al.* 2009a. Large bio-geographical shifts in the north-eastern Atlantic Ocean: from the subpolar gyre, via plankton, to blue whiting and pilot whales. Progress in Oceanography, 80: 149–162.

Hátún, H., Payne, M. R., and Jacobsen, J. A. 2009b. The North Atlantic subpolar gyre regulates the spawning distribution of blue whiting (*Micromesistius poutassou*). Canadian Journal of Fisheries and Aquatic Sciences, 66: 759–770.

Hilborn, R., and Ovando, D. 2014. Reflections on the success of traditional fisheries management. ICES Journal of Marine Science, 71: 1040–1046.

Hilborn, R., and Walters, C. 1992. Quantitative Fisheries Stock Assessment: Choice, Dynamics, and Uncertainty. Chapman & Hall, New York. 570 pp.

Hurtado-Ferro, F., Szuwalski, C. S., Valero, J. L., Anderson, S. C., Cunningham, C. J., Johnson, K. F., Licandeo, R., *et al.* 2014. Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. ICES Journal of Marine Science, 72: 99–110.

ICES. 2012. Report of the Workshop on Frequency of Assessments (WKFREQ), 6–8 March 2012, By Correspondence. ICES Document CM 2012/ACOM: 34. 29 pp.

ICES. 2013. Report of the Blue Whiting/Norwegian Spring-Spawning (Atlanto-Scandian) Herring Workshop (WKBWNSSH), 11–13 March 2013, Bergen, Norway. ICES Document CM 2013/ACOM: 69. 88 pp.

ICES. 2014. Report of the Working Group on Widely Distributed Stocks (WGWIDE), 26 August–1 September 2014, ICES Headquarters, Copenhagen, Denmark. ICES Document CM 2014/ACOM:15. 938 pp.

Kell, L., Pilling, G., Kirkwood, G., Pastoors, M., Mesnil, B., Korsbrekke, K., Abaunza, P., *et al.* 2006. An evaluation of multiannual management strategies for ICES roundfish stocks. ICES Journal of Marine Science, 63: 12–24.

Kelly, C. J., and Codling, E. A. 2006. 'Cheap and dirty' fisheries science and management in the North Atlantic. Fisheries Research, 79: 233–238.

Langøy, H., Nøttestad, L., Skaret, G., Broms, C., and Fernö, A. 2012. Overlap in distribution and diets of Atlantic mackerel (*Scomber scombrus*), Norwegian spring-spawning herring (*Clupea harengus*) and blue whiting (*Micromesistius poutassou*) in the Norwegian Sea during late summer. Marine Biology Research, 8: 442–460.

Marchal, P. 1997. Managing growth overfishing with multiannual compromise strategies. Canadian Journal of Fisheries and Aquatic Sciences, 54: 2255–2276.

Marchal, P., and Horwood, J. 1995. Multi-annual TACs and minimum biological levels. ICES Journal of Marine Science, 52: 797–807.

Maunder, M. N., and Punt, A. E. 2013. A review of integrated analysis in fisheries stock assessment. Fisheries Research, 142: 61–74.

Myrseth, J., Enberg, K., Heino, M., and Fiksen, Ø. 2011. Do accurate stock estimates increase harvest and reduce variability in fisheries yields? Natural Resource Modeling, 24: 222–241.

Nielsen, A., and Berg, C. W. 2014. Estimation of time-varying selectivity in stock assessments using state-space models. Fisheries Research, 158: 96–101.

Nøttestad, L., Krafft, B. A., Anthonypillai, V., Bernasconi, M., Langård, L., Mørk, H. L., and Fernö, A. 2015. Recent changes in distribution and relative abundance of cetaceans in the Norwegian Sea and their relationship with potential prey. Frontiers in Ecology and Evolution, 2: 83.

Patterson, K., Cook, R., Darby, C., Gavaris, S., Kell, L., Lewy, P., Mesnil, B., *et al.* 2001. Estimating uncertainty in fish stock assessment and forecasting. Fish and Fisheries, 2: 125–157.

Payne, M. R., Egan, A., Fässler, S. M. M., Hátún, H., Holst, J. C., Jacobsen, J. A., Slotte, A., *et al.* 2012. The rise and fall of the NE Atlantic blue whiting (*Micromesistius poutassou*). Marine Biology Research, 8: 475–487.

Prokopchuk, I., and Sentyabov, E. 2006. Diets of herring, mackerel, and blue whiting in the Norwegian Sea in relation to *Calanus finmarchicus* distribution and temperature conditions. ICES Journal of Marine Science, 63: 117–127.

Punt, A. E. 2015. Strategic management decision-making in a complex world: quantifying, understanding, and using trade-offs. ICES Journal of Marine Science, doi:10.1093/icesjms/fsv193.

Rochet, M.-J., and Rice, J. C. 2009. Simulation-based management strategy evaluation: ignorance disguised as mathematics? ICES Journal of Marine Science, 66: 754–762.

Sætre, R., Toresen, R., and Anker-Nilssen, T. 2002. Factors affecting the recruitment variability of the Norwegian spring-spawning herring (*Clupea harengus* L.). ICES Journal of Marine Science, 59: 725–736.

Skagseth, Ø., Slotte, A., Stenevik, E. K., and Nash, R. D. 2015. Characteristics of the Norwegian Coastal Current during years with high recruitment of Norwegian spring spawning herring (*Clupea harengus* L.). PLoS One, 10: e0144117.

Vikebø, F. B., Korosov, A., Stenevik, E. K., Husebø, Å, and Slotte, A. 2012. Spatio-temporal overlap of hatching in Norwegian spring-spawning herring and the spring phytoplankton bloom at available spawning substrata. ICES Journal of Marine Science, 69: 1298–1302.

Handling editor: Emory Anderson