Ecological Indicators 156 (2023) 111084

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Assessing change in the occurrence of rare species using the binomial distribution

W. Nikolaus Probst [a,*], Christopher P. Lynam [b], Joanna K. Bluemel [c], Maurice Clarke [d]

[a] Thünen-Institute of Sea Fisheries, Herwigstraße 31, 27572 Bremerhaven, Germany
[b] Centre for Environment, Fisheries and Aquaculture Science, Lowestoft Laboratory, Pakefield Road, Lowestoft, Suffolk NR33 0HT, UK
[c] Centre for Environment, Fisheries and Aquaculture Science, The Nothe, Barrack Road, Weymouth, Dorset DT4 8UB, UK
[d] Marine Institute, Oranmore, Galway, Ireland

A R T I C L E I N F O

Keywords:
Population status
Assessment
North Atlantic
Sensitive fish species
OSPAR
Binomial distribution

A B S T R A C T

The demand for comprehensive biodiversity assessments is increasing through the implementation of the ecosystem-based approach to management (EBM) of marine resources. Regional sea conventions such as the Oslo-Paris Commission (OSPAR) strive to implement EBM by developing an extensive status assessment program for the marine environment. Demersal fish communities are one ecosystem feature assessed by OSPAR through several ecological indicators. One of these indicators assesses the recovery in the population abundance of sensitive fish species, which was initially developed to report on the status of the sensitive fish community as a whole. However, for national reporting obligations, contracting parties of OSPAR (particularly for EU member states reporting to the Marine Strategy Framework Directive) prefer to have assessments for individual species. The previous indicator assessment relied on a suite of demersal fish species, which were caught frequently in scientific groundfish surveys, but did not provide assessment results for rare species caught in low frequencies. This study introduces a new assessment approach, the Binomial Occurrence Assessment (BOA), for the FC-1 indicator now renamed the "Recovery of sensitive fish species", by applying the binomial distribution to relative occurrence data from scientific fisheries surveys. BOA uses occurrences in a reference period to determine boundaries for the expected occurrences in the recent (six year) assessment period of each survey. Significant changes in occurrence between the reference and assessment period, i.e. declines or recoveries, can then be detected when the observed occurrences in the assessment period fall outside of these boundaries. Methods to integrate the assessment results across multiple surveys are explored and compared since data on occurrences for fish species are available from more than one fisheries survey in each marine region considered by OSPAR. A case study on the sensitive demersal fish species in the North Sea exemplifies the applicability of BOA. Furthermore, assessments by BOA are compared to data from analytical stock assessments for those data-rich sensitive species that can support both approaches. Despite some shortcomings of BOA, such as the inability to detect declines of very rare species and the potential for occurrence metrics to differ from abundance metrics, the BOA allows an assessment of the status of a wide suite of fish species throughout the entire OSPAR region. The low data requirements of BOA allow its generic application to any other monitoring program that has captured occurrences of single species or species suites in the past and present.

1. Introduction

The increasing human population of the world exerts an ever-increasing pressure on marine ecosystems at both local and global scales (Halpern et al., 2008; Halpern et al., 2012; Korpinen et al., 2021). Impacts on marine ecosystems arise from a wide range of activities including fishing, shipping, aquaculture, agriculture, aggregate extraction, wastewater treatment, waste disposal, tourism, oil exploration, offshore wind power generation and other renewable energy developments (Curtin and Prellezo, 2010; Borja et al., 2016; Stelzenmüller et al., 2022). Therefore, governance authorities worldwide are striving to develop ecosystem-based management plans for human activities to minimise impacts on marine life and achieve the sustainable use of marine resources (Grumbine, 1994; Arkema et al., 2006; Curtin and

* Corresponding author.
E-mail address: nikolaus.probst@thuenen.de (W. Nikolaus Probst).

https://doi.org/10.1016/j.ecolind.2023.111084
Received 16 July 2023; Received in revised form 9 October 2023; Accepted 11 October 2023

Prellezo, 2010; Katsanevakis et al., 2011; Link and Browman, 2014).

Ecosystem-based management aims to consider the combined impacts of human activities on marine life, and therefore encompasses assessments on the status of multiple ecosystem features such as marine fish, birds and mammals (Levin et al., 2009; Greenstreet et al., 2012; Harvey et al., 2017). In Europe, the EU's Marine Strategy Framework Directive (MSFD) and the Regional Seas Conventions (e.g. the Convention for the Protection of the Marine Environment of the North-East Atlantic also known as the Oslo-Paris Commission, OSPAR, and The Baltic Marine Environment Protection Commission also known as the Helsinki Commission, HELCOM) have triggered the development of new assessment programs with new ecological indicators aiming to assess the status of different ecosystem features (Greenstreet et al., 2011; Greenstreet et al., 2012; HELCOM, 2013; OSPAR, 2017).

Indicators developed for biodiversity assessments often rely on data from existing monitoring programmes originally designed for other purposes. Scientific groundfish surveys have originally been designed to meet the data requirements of analytical stock assessments focusing on commercially important target species. However, in these surveys a wide array of additional species are caught and recorded. Given that monitoring programmes have not typically been initiated or adapted to the broader aims of EBM, novel methods that can use existing data appropriately are essential to fulfil the increasing assessment requirements of marine policies with limited resources at hand (Edgar et al., 2016).

In this study, we present a novel approach to detect the decline or recovery in populations of species based on occurrences in scientific survey data. We assess change in the frequency of occurrence of populations between two time-periods using the binomial distribution to test for significant deviations in an assessment period versus expected occurrences observed in a reference period. We illustrate this novel approach for selected demersal fish species in the North Sea, which are sensitive to anthropogenically induced mortality (Greenstreet et al., 2012; Gascuel et al., 2014). Furthermore, we propose and contrast methods to integrate the outcomes of assessments when multiple datasets exist for the same population. This approach was the basis for the assessments in the OSPAR Quality Status Report 2023 for the indicator "Recovery of sensitive fish species" (known as Fish and Cephalopod indicator number one, FC-1). This methodology was based on an earlier version of FC-1 developed by Greenstreet et al. (2012), who used time series of abundance of demersal fish in scientific fisheries surveys to provide an overall status assessment for the sensitive fish component of the demersal fish community. A key aim of the indicator is to identify those populations of sensitive species that are not assessed elsewhere (e.g. by fish stock assessment) and are potential management concerns (i.e. declining).

2. The binomial occurrence assessment (BOA)

2.1. Methodological development of FC-1 indicator "Recovery of sensitive fish species"

The OSPAR indicator FC-1 was derived from the original work of Greenstreet et al. (2012), which focused on a suite of demersal fish species considered to be sensitive to anthropogenically induced mortality. The approach by Greenstreet et al. (2012) was based on the time series of catch-rates per species (abundance per hour data). Only species that occurred in at least 50 % of all the years, by the International Bottom Trawl Survey in Quarter 1, were included in the suite of assessed species. The temporal changes in this suite of selected species were considered to be representative for the entire sensitive demersal fish community. After assessing the status of single species, Greenstreet et al. (2012) employed the binomial distribution, with a probability of success from a random walk model, to determine the number of species that would by chance achieve a species-specific assessment target, i.e. a catch rate at or above the upper 25th percentile of all annual catch rates. This approach was used to derive an assessment threshold to test whether significantly more species were increasing in abundance than would be expected by chance and hence whether the sensitive fish community as a whole could be considered to be recovering.

The OSPAR Intermediate Assessment in 2017 (IA2017) implemented the approach by Greenstreet et al. (2012) for multiple surveys in multiple regions. The integrated outcomes across multiple species in each survey were tested to identify whether recovery was occurring per survey. The outcomes (recovery overall or not) for multiple surveys were then integrated also by the application of the binomial distribution to provide a single outcome per OSPAR Region. However, the indicator was criticised for not providing status information per species, i.e. assessment results were presented following aggregation for the sensitive fish community per survey and overall per region. Following the IA2017, several EU member states requested species status assessments for the use in their national reporting for the MSFD. Consequently, OSPAR requested that the assessment protocol of FC-1 was further developed by ICES who commissioned a workshop on the production of abundance estimates for sensitive species (WKABSENS) to provide information on individual species and to include previously omitted species where data allowed (ICES, 2021).

The new Binomial Occurrence Assessment (BOA) for sensitive fish populations conceived within WKABSENS (ICES, 2021) is similar to the original FC-1 by OSPAR (2017) and Greenstreet et al. (2012) in that it utilises the binomial distribution to contrast data and test for significant deviations from expected values. Whereas Greenstreet et al. (2012) used this technique to integrate species outcomes from time series data within a single fisheries survey, BOA uses the approach to contrast the probability of occurrence of species between two assessment periods. BOA can distinguish between significant declining, recovering or stable occurrence patterns and it is also appropriate for rare and data-poor species for which other methods, such as time series-based assessments of abundance, may not be applicable (Gröger et al., 2011; Greenstreet et al., 2012; Lindegren et al., 2012; Probst and Stelzenmüller, 2015).

2.2. The new FC1-approach to identify population status changes

To allow for the assessment of populations from a wide array of species, including those that are rare and/or caught irregularly by surveys and in low frequencies, we developed a new assessment methodology that compares the number of positive occurrences in an assessment period to the number of occurrences in a previous reference period and correcting for changes in sample size. The limit specified by Greenstreet et al (2012) and subsequently OSPAR (2017) that each species must be detected in at least 50 % of years in the survey to provide reliable catch-rates of abundance is relaxed within BOA to a minimum threshold of 5 occurrences of each species in the entire survey time series. This simple comparison provides all relevant parameters for the binomial distribution, including the probability of capture (i.e., frequency of occurrence) during the reference period and the known number of samples taken during the assessment period and the known number of successes (positive occurrences in the assessment period). Given chosen probability levels (i.e. significance), the associated quantiles from the cumulative distribution function can be considered as thresholds for occurrence in the assessment period that indicate significant increases or decreases in the occurrence of a species in the assessment period (Fig. 1).

The binomial model predicts the probability of n successful outcomes of a Bernoulli experiment that has two possible outcomes (e.g. the toss of a coin). The binomial distribution then gives the probability of k successes in n trials of the experiment with a fixed probability of the single success p (e.g. heads on a coin $p = 0.5$ or rolling a six on a dice $p = 0.167$):

$$P(k|n, p) = \binom{n}{k} p^k (1 - \mathrm{p})^{n-k} \quad (1)$$

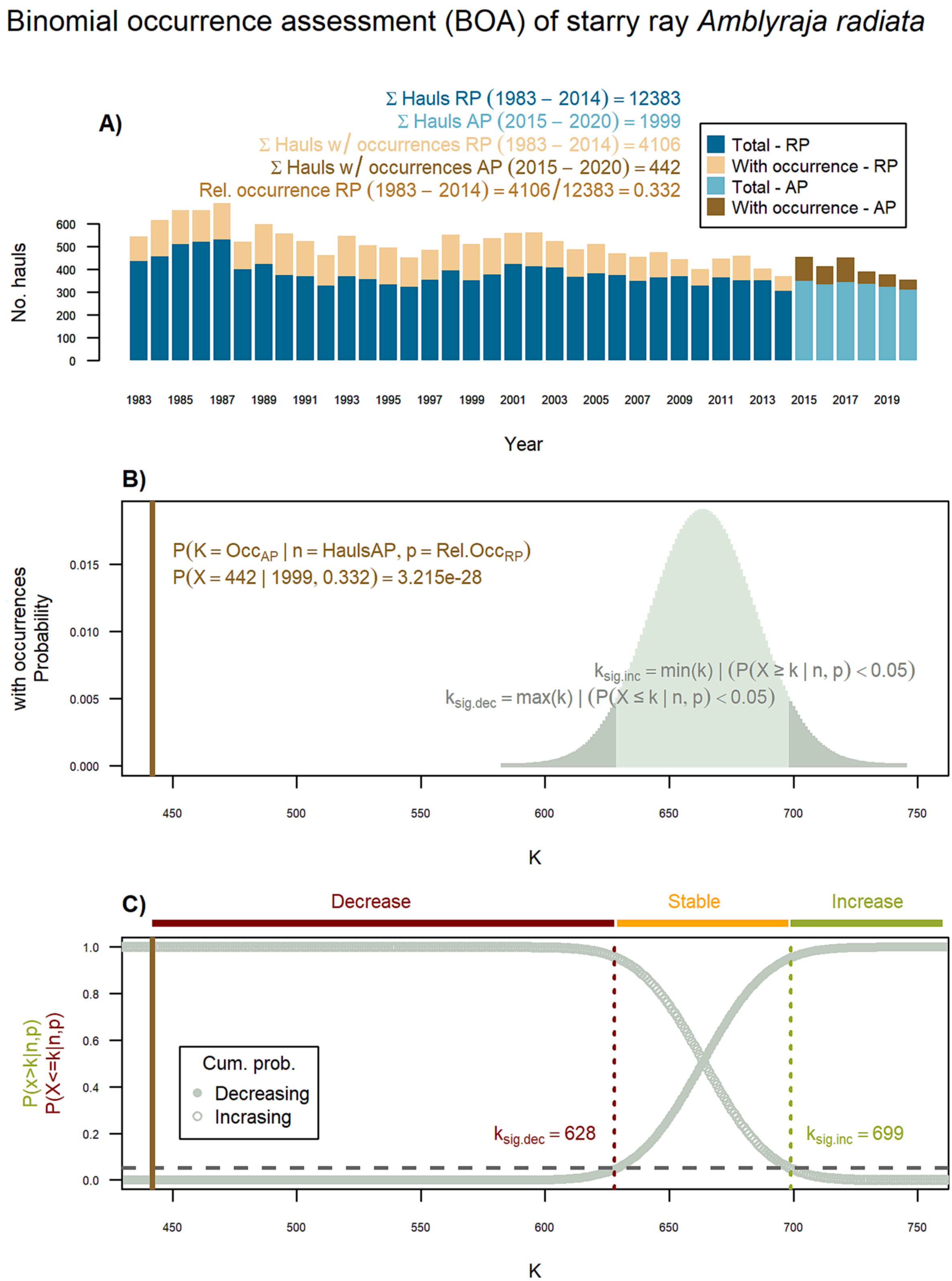

Fig. 1. The concept of using the binomial occurrence assessment (BOA) is exemplified for starry *ray Amblyraja radiata* in the North Sea. **A)** Time series of the total number of hauls and hauls with occurrences from 1983 until 2020. The time series is split into a reference (RP) and an assessment period (AP) to determine the number of total hauls and hauls with occurrences in both periods. These figures allow to calculate the parameters p (relative occurrence in RP), k (number of hauls with occurrences in AP) and n (total number of hauls in AP). **B)** Inserting p, k and n into the binomial distribution provides the probability distribution with lower and upper tails to determine threshold values ($k_{sig.dec}$ and $k_{sig.inc}$). The probability of observing $k = 442$ hauls with occurrences in a sample size of $n = 1999$ is thereby $\ll 0.001$. **C)** The cumulative probability functions indicate at which k the probability falls below the significance threshold of 0.05 (k_{sig}). Observing 628 or fewer hauls with occurrences in AP thereby indicates a significant decline, observing 699 or more hauls with occurences indicates a significant increase. Observing between 629 and 698 occurrences in the AP would indicate no increase and no decline, and therefore, would represent a stable situation.

The cumulative distribution function of the binomial distribution determines the probability of k or fewer successes:

$$P(X \leq k|n, p) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i} \tag{2}$$

Using the cumulative distribution function, it is possible to determine the values of k for which Eq.2 is below a predefined significance threshold, e.g. $\alpha < 0.05$. These values of k represent the lower tail of the binomial distribution (Fig. 1B) and any observed k in this tail would indicate a significant deviation from an expected mean. Hence, the largest k value for which Eq.2 is $< \alpha$ can be used as a threshold K_{sig} to identify the significant deviation from the expected mean. Thus, when the number of occurrences in n hauls within the assessment period is equal or less than the maximum k required to satisfy the condition $P(X \leq k|n, p) < \alpha$, a significant decline in occurrences relative to the reference period can be assumed. In the example given in Fig. 1, the number of hauls in the assessment period is $n = 1999$ and the expected probability of success is 0.332 (equal to the relative occurrence in the reference period). Hence the maximum k is 628, before the significance of the test increases above the threshold level of $\alpha = 0.05$. Accordingly, any number of occurrences up to and including 628 indicate a significant decrease, which is the case in the example with 442 occurrences observed in the reference period.

The binomial distribution can be used to estimate the probability of observing k occurrences of a species in n samples (here fishing hauls) in a survey in a particular assessment period once we have an estimate of the probability p of detecting the species in a single haul. The key assumptions here are that each haul in the survey data is considered as an independent Bernoulli-experiment and the probability p of detecting the species remains constant in the reference and assessment periods. Based on these assumptions, p can be estimated from the frequency of occurrence of the species in the survey in a chosen reference period, where the frequency of occurrence is simply the number of hauls with occurrence divided by the total number of hauls. According to Eq.2, a threshold $k_{sig.}$ can be set where any observed k in the assessment period becomes significantly unlikely, and thus indicates a statistically relevant decline in occurrence when compared to an expected occurrence derived from the reference period.

The counter-event for the upper tail of the binomial distribution can be used accordingly to set a threshold for indicating a statistically significant increase, i.e. recovery (where the probability is below the predefined significance threshold) in the species' occurrence in the assessment period as follows:

$$P(X > k|n, p) = 1 - P(X \leq k|n, p) = 1 - \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i} \tag{3}$$

and

$$P(X \geq k|n, p) = P(X > (k-1)|n, p) = 1 - \sum_{i=0}^{k-1} \binom{n}{i} p^i (1-p)^{n-i} \tag{4}$$

Hence, a significant recovery in occurrences can be assumed, relative to the reference period for which p was set, if the number of occurrences in n hauls within the assessment period is equal to or greater than the minimum k required to satisfy the condition: $P(X \geq k|n, p) < \alpha$.

2.3. *Long-term and short-term assessments for FC-1 indicator "Recovery of sensitive fish species"*

Information on both long-term and short-term changes in populations can be determined from the same dataset by changing the length of the reference period (RP) but maintaining the same assessment period (AP). To align with the 6-year cycle of the MSFD, the AP is set as the most recent six years available in the survey data, e.g. Table S1). A long-term reference period (RLP) is set as the starting year for each survey until the year prior to the assessment period, while the short-term reference period (RPS) is the six years immediately prior to the assessment period.

The status of a species within a survey was labelled "recovering" if $k \geq k_{sig.inc}$, "decreasing" if $k \leq k_{sig.dec}$, and "stable" if $k \geq k_{sig.dec}$ and $\leq k_{sig.inc}$. The species' status was assessed as "unknown" if $k_{sig.dec}$ was zero or if the total number of occurrences in the full time series was less than five.

3. Combining assessments from multiple surveys for species status

In many OSPAR regions fish species occurrence data are available from more than one scientific survey. To include all information and consider each survey, we developed and compared two integration approaches and one selection approach. The first integration approach is based again on the binomial distribution, the second integration approach is based on a combined index of survey suitability based on the relative occurrence in the reference period and frequencies of BOA outcomes from the different surveys (i-score). The selection approach based the assessment on whichever survey had the highest relative occurrence in the reference period (termed the "best survey" for that species in that region).

3.1. *Binomial integration [BI]*

The significance level (α) chosen to identify significant change in occurrence of a species was chosen as 0.05. The binomial integration (BI) approach considers the probability of a type 1 error in order to combine multiple outcomes from multiple surveys of a species. Using n as the number of surveys in a region and k as the number of surveys with significant recoveries or declines, the probability of observing recoveries in k out of n surveys can be determined according to Eq. (2) and Eq. (4). If, for example, in one region eight surveys were operating ($n = 8$) and three surveys indicated a significant recovery ($k = 3$), the probability for $k = 3$ is:

$$P(k = 3|8, 0.05) = \binom{3}{8} 0.05^3 0.95^5$$

$$P(k = 3|8, 0.05) = 0.00541$$

Hence, it is significantly improbable ($P < 0.05$) that three surveys would each indicate a recovery by chance, suggesting that the observed frequency of recoveries indicates a real recovery in the occurrence of the species. However, at the same time, three other surveys out of the total of eight might indicate a significant decline, which is also a significant deviation from a randomly expectable number of declines. Consequently, the binomial integration approach can result in a mixed integration outcome, because significant numbers of surveys can simultaneously indicate a recovery and decline.

3.2. *I-score [ISCR]*

To overcome the problem of a potentially mixed integration result using BI, a second integration approach, the i-score (ISCR), was developed. The i-score is an index that combines the relative frequencies of assessment outcomes (RF_{AO}) with the average relative occurrence in the reference period ($\overline{Occ_{RP}}$) as a product of both metrics (Table 1):

$$i.score = RF_{AO} * \overline{Occ_{RP}}$$

The i-score thereby selects for the most frequently observed assessment outcome combined with the highest average relative occurrence, assuming that these surveys have the best representative data to assess the state of the species.

Table 1
Examples of integrating or selecting BOA results of single surveys using the binomial survey integration [BI], the best survey [BS] and the i-score [ISCR] approaches. RF_{AO} = relative frequency of BOA assessment outcomes; $\overline{Occ}_{RP}$ = average frequency of occurrence in the reference period. Surveys with the highest relative occurrence in the reference period, i.e. 'best' (S5 in scenario 1, S4 in scenario 2) are highlighted in bold. The BOA outcomes of these 'best' surveys provide the assessment outcomes of the best survey approach. BI = values of binomial integration derived from Eq. (2). Integration results are symbolised as ↑=recovery, ↓= decline, ↕=mixed, ↔=stable.

Survey	BOA result	Rel. occurrence in RP	Freq. BOA outcomes	RF_{AO}	$\overline{Occ}_{RP}$	i-score [ISCR]	Binomial integration [BI]	BI result	ISCR result	BS result
Case 1: Apparent decrease scenario										
S1	Recovering	0.01	1	0.125	0.01	0.001	0.280	↓	↓	↓
S2	Stable	0.33	3	0.375	0.183	0.069	NA			
S3	Stable	0.21								
S4	Stable	0.01								
S5	Declining	**0.45 (BS)**	4	0.500	0.205	**0.103 (highest i-score)**	**< 0.001**			
S6	Declining	0.12								
S7	Declining	0.07								
S8	Declining	0.18								
Case 2: Mixed scenario										
S1	Recovering	0.01	3	0.375	0.183	**0.069 (highest i-score)**	**0.005**	↕	↑	↔
S2	Recovering	0.33								
S3	Recovering	0.21								
S4	Stable	**0.45 (BS)**	2	0.250	0.230	0.058	NA			
S5	Stable	0.01								
S6	Declining	0.12	3	0.375	0.123	0.046	**0.005**			
S7	Declining	0.07								
S8	Declining	0.18								

3.3. *Best survey [BS]*

The third alternative approach that accounts for the existence of multiple potential assessment outcomes from multiple surveys, selects the 'best' survey (BS) based on the relative occurrence in the reference period (Table 1). The 'best' survey thus represents the survey with the highest frequency of occurrence with the underlying assumption that this survey thereby provides the most reliable information on change in the occurrence of the species.

4. Data and methods for case studies

4.1. *Sensitive fish in the North Sea*

BOA was applied to 38 sensitive fish populations in the North Sea, with the selection and both integration approaches, using data from eight surveys (Fig. 2). The species were a subset of those identified as sensitive by ICES (2021) for which data were available in the survey data product (Table S1) prepared by Lynam and Ribeiro (2022a) using data extracted from the ICES Database of Trawl Surveys https://datras.ices.dk).

No assessment was possible for survey CSEngBT1 due to the short length of the available time series (four years). For three surveys, BBIC(n)SpaOT4, CSNIrOT4 and BBICFraBT4, no RPL was available due to the length of the time series (12 years or less). For BBIC(n)SpaOT4 and BBICFraBT4 (length of eight and ten years, respectively) the AP and RPS were each reduced to four or five years, respectively.

4.2. *Comparison of BOA against analytical stock assessments*

Although the FC-1 indicator "Recovery of sensitive fish species" aims to identify those populations of sensitive species not assessed elsewhere that are potential management concerns (i.e. declining status), a subset of the sensitive species identified by ICES (2021) also support commercially fished populations in some areas (stocks). To support commercial fisheries, fish stocks are typically of relatively high abundance and the target of scientific surveys and thus often considered data-rich species suitable for analytical fish stock assessment by ICES.

For the subset of the sensitive species with survey data and where full analytical stock assessments have been made (ICES, 2022a), the surveys were filtered to include only those that are relevant to the fish stock area (i.e. include the stock area in part or fully). Furthermore, to improve comparability with the spawning stock biomass (SSB) metric, surveys were retained in the analysis only if the mean length of individuals (weighted by biomass) of the species, using data prepared by Lynam and Ribeiro (2022b), was greater than the mean length of maturity for the species (Table S2). This filtering led to eight species with multiple survey datasets and 14 stock assessments for which comparisons were considered relevant (Table S3). Although changes in SSB and occurrence need not show similar patterns due to a range of ecological processes impacting the spatial distribution of species and/or biomass, including climate-forced distribution shifts (see Link et al. 2011), it is unlikely that they would directly contradict each other in the majority of cases (i.e. increase vs decrease).

Changes in SSB, between assessment and reference period, of each of these 14 stocks was compared against the outcomes of the three different BOA integration methods (BS, BI and ISCR). For direct comparability, the same reference period (2009–2014) and assessment period (2015–2020) was used for each comparison. The difference in mean SSB in the reference and assessment period was assessed statistically using a *t*-test. If the *t*-test indicated a significant difference between both means, the change in SSB was classified as "declining" or as "recovering". In case of a non-significant *t*-test result, SSB was classified as "stable".

When comparing SSB and BOA outcomes, they were classified as "agreement" if both procedures resulted in the same classification and show "contradiction" if one procedure suggested a decline while the other an increase. If one procedure suggested a stable or mixed outcome and the other did not the outcomes were said to show "divergence".

5. Results and discussion

5.1. *Sensitive fish in the North Sea*

Comparing the binomial, best survey and I-score integration, all methods yielded equal assessment outcomes for 21 species (55 %) in the long-term assessment and 24 species (63 %) in the short-term assessment. The highest number of recoveries was found for the RPS and the BI integration (17 out 38 species), the lowest number of recoveries was found for RPS and BS (11out of 38 species). Starry ray *Amblyraja radiata*, eel *Anguilla anguilla*, lump sucker *Cylcopterus lumpus*, sandy ray *Leucoraja circularis*, pollack *Pollachius pollachius* and eelpout *Zoarces viviparous* indicated declining occurrences in at least two integration/selection

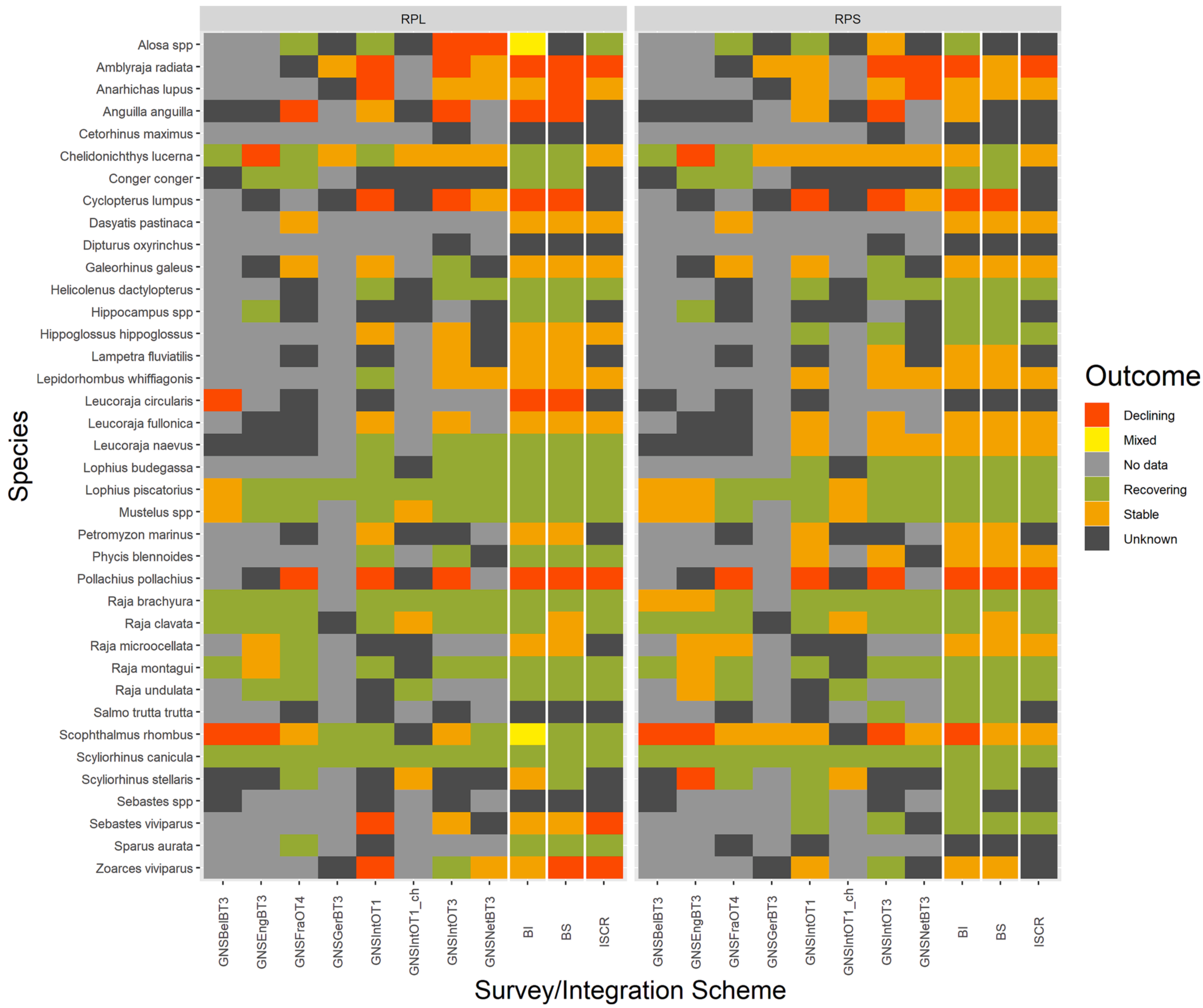

Fig. 2. Assessment and integration outcomes for 38 fish species of the Greater North Sea. RPL = long reference period (full time series), RPS = short reference period (six years previous to assessment period), BI = binomial integration, BS = best survey, ISCR = i-Score. For a description of survey abbreviations refer to supplementary Table S1.

methods in the RPL, suggesting long-term declines of these species in the Greater North Sea. Consistent long-term recoveries were indicated for 11 species. Rare or rarely caught species such as basking shark *Cetorhinus maximus*, long-nosed skate *Dipturus oxyrhinchus* or Sebastes spp. (with fewer than 5 occurrences) could not be assessed.

5.2. Comparison of BOA against analytical stock assessments

From the three integration approaches of BOA, the ‘Best survey’ approach showed the highest level of agreement with the analysis of changes in SSB (10 out 14, i.e. 71 %, Fig. 3). Contrary, the ‘binomial integration’ and the ‘I-score’ had a lower frequency of agreement (7 out of 14, 50 %). Contradictions between changes in SSB and BOA were only observed for two monkfish stocks (mon.27.78abd and mon.27.8c9a) accounting for not more than 14 % with any of integration or selection approaches trialled. If the procedures led to contradictions at random, we would expect 22 % of comparisons to contradict directly for BS and ISCR and 17 % for the BI. Therefore, the analysis suggests that BOA should lead to non-contradictive or equal assessment results relative to change in SSB in the majority of cases.

Divergence between BOA and SSB-SA is also seen for 6 stocks (Fig. 3B, two of which diverge in each of the BOA approaches (i.e. cod27.7a and meg27.8c9a). Cod in the Irish Sea (cod27.7a) was found to be stable in terms of SSB but it was declining in occurrence in two (CSNIrOT1[BS] and CSNIrOT4) otter trawl surveys (Fig. S2.1 & S2.2). The beam trawl (CSEngBT3) survey for Irish Sea cod was the only one that sampled the entire sea area, and was found to be stable, but this was not considered the Best Survey due to low occurrences in the samples. For this stock the survey data are highly variable (Fig. S2.1). In contrast, megrim (meg27.8c9a) is showing consistent increase in all four surveys that sample this species, but the applied *t*-test did not confirm the significance of the positive trend in SSB since 2018.

The BI outcomes for Celtic Sea/Bay of Biscay megrim stock (meg27.7b-k8adb) diverges from BS and ISCR due to the mixed outcome (with 2 surveys increasing and two decreasing), whereas the BS and the ISCR agree, showing the intended benefit of the ISCR method. In contrast, outcomes for cod to the West of Scotland (cod27.6a) for SSB and BI agree that the stock is stable, but BS and ISCR (weighted toward the BS) indicate a recovery, demonstrating the weakness of the reliance on a single survey.

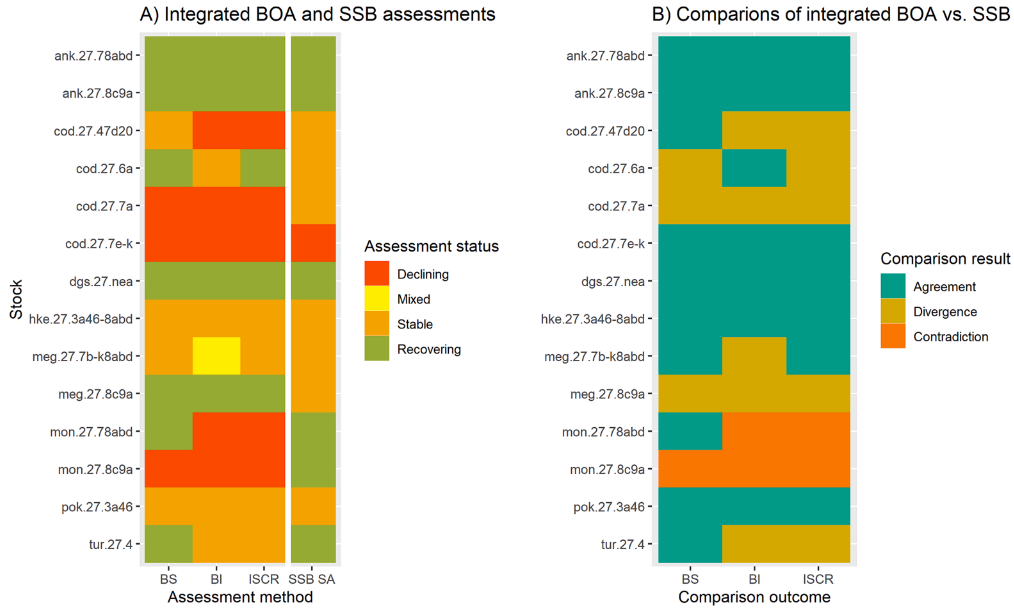

Fig. 3. Analysis of BOA assessments and trends in spawning stock biomass (SSB SA) from analytical stock assessments for 14 stocks of various OSPAR regions. **A)** Assessment results of integrated BOA by three approaches and trend assessment of SSB SA. **B)** Comparisons of integrated BOA methods vs. trend assessment of SSB SA. BOA integration approaches are abbreviated as BS ='Best survey', BI ='Binomial integration' and ISCR ='I-score'.

Cod in the North Sea (cod27.47d20) is stable in ICES area 4 during winter according to the BS (GNSIntOT1), which agrees with the SSB outcome. However, a divergent trend (decreasing) is observed in summer (GNSIntOT3) and in two surveys in the eastern channel (GNSIntOT1_channel and GNSFraOT1). This suggests the generic analysis of changes in SSB- may not capture seasonal and spatial changes within the stock area as discussed further by ICES (2022b).

The North Sea turbot (tur.27.4) outcome, based on the Dutch beam trawl survey (GNSNetBT3) with wide spatial coverage (and BS), agrees with SSB-SA. However, the German (GNSGerBT3) and Belgian (GNSBelBT3) beam trawl survey results diverge from SSB, which is likely due to the limited area these surveys cover (central-eastern and south-western North Sea respectively). In contrast, the otter trawl surveys are spatially extensive with highly variable frequency of occurrence and potentially not representative of stock level changes.

6. Advantages, caveats and limits of the BOA

6.1. Advantages of the BOA

6.1.1. Generic application

The major advantage of the BOA is its easy implementation using already existing data, e.g. from scientific fisheries surveys used in our study. Thus, the BOA approach provides a status assessment according to the requirements of OSPAR and the MSFD (in particular the MSFD criterion D1C2 – abundance of species) and closes knowledge gaps on previously unassessed ecosystem features.

The development of BOA was based on scientific fisheries surveys on demersal fish in the Northeast Atlantic, but BOA can be applied to any dataset with presence-absence data collected in the past and present. Such datasets could relate to birds, marine mammals, insects or plants. Examples of potential data sources can be found at the Global Biodiversity Information Facility (https://www.gbif.org, e.g. for butterflies and bumble bees in Norway; Åström and Åström, 2022). BOA is not limited to demersal fish nor groundfish survey data, and it may not even require a regular temporal sampling scheme. BOA can be applied as long as there are sufficient data to estimate the probability of occurrence in a reference period (e.g. by an irregular sampling scheme in several years) and comparable sampling was conducted in an assessment period. The latter point on 'comparable sampling' is important since species identification even within scientific surveys can improve/deteriorate over time and the ability of the fishers to operate their fishing gear can also change. We have used a standardised dataset (Lynam and Ribeiro, 2022a) that accounts for known issues (due to change in methodology and reporting) and does not include the data for the North Sea surveys in the 1960 and 1970s for this reason (incomparable sampling) following Moriarty et al. (2017). Furthermore, the application of BOA is limited to data sets containing true absence information, as presence-only data will not allow determination of the relative occurrence of species, and hence no values for p can be obtained. Although, by definition, rare species have low catchability in trawl surveys, the BOA approach compares changes in occurrence between periods (that have equal low catchability) so this standardisation reduces the impact of low catchability on the assessment.

6.1.2. Easy implementable mathematics

BOA does not require extensive knowledge on modelling techniques or excessive computing power. By contrast, sophisticated statistical models, e.g. to estimate and model abundance, can become very demanding on computer resources and expert knowledge (Thorson et al., 2017).

Essentially, BOA is implementable with widely distributed or freeware software such as spreadsheets or the R programming language (R Core Team, 2013). An example for a one-line code implementation in R is given in supplements S3.

6.1.3. Insensitivity to gaps in time series

BOA is rather insensitive to gaps in time series. Data from a reference period can be pooled from several years to obtain an estimate of relative occurrences in the reference period. Thereby BOA can be applied to irregular monitoring schemes sampling data at irregular intervals. Although, comparable sampling in the reference and assessment periods is advised to ensure equal chance of detecting occurrences in both time periods.

6.2. *Limitations of the BOA*

6.2.1. *Detectability of declines*

There are limitations to BOA for assessing the significance of a decline in cases where *p* and/or *n* are low (Fig. 4). In these circumstances the cumulative distribution function does not approach probabilities below 0.05 and hence $k_{sig.dec}$ cannot be determined. In other words, due to low sample sizes and/or relative occurrences (i.e. catch rates) the probability of not encountering a species at least once is rather probable ($p > 0.05$).

By contrast, the probability of an increase can always be significantly determined (if zero occurrence is the norm, a single occurrence already can indicate a significant increase). This finding may have implications for the assessment target, such as suggested by Greenstreet et al. (2012) or Probst and Stelzenmüller (2015), i.e. whether to use the assessment against a significant decline or recovery in determining the status of a species. If assessed against a decline, assessments may not be possible for very rare species, whereas an assessment against a recovery technically will always be possible. However, it should be noted that in the case study of the Greater North Sea (as in all other OSPAR areas) a minimum of at least five occurrences in the survey time series was required to implement the BOA, otherwise the status of the species was classified as "unknown".

6.2.2. *Autocorrelation*

The BOA assumes that the probability of occurrence *p* (i.e. the relative occurrence in the reference period) is constant. However, this may be not the case, as the probability of occurrence *p* for a species may differ between sampling sites. For example, it may be more likely to catch a certain fish species on sandy than on muddy habitats. Hence, within a fisheries survey, *p* will most likely be unequal among single hauls. The BOA approaches this problem by estimating an average *p* through the relative occurrence during the reference period assuming that the survey design is unchanged between assessment and reference periods.

The sensitivity of BOA against non-constant values of *p* can be tested by applying Fast Fourier convolution (FFC). Given a vector of differing $p = (p_1, p_2, \ldots, p_i)$, $k_{sig.FFT}$ can be calculated using FFC convolution. This $k_{sig.FFC}$ can then be compared against boundaries derived from binomial distribution ($k_{sig.binom}$) with *p* as the average of $p_1 - p_i$.

In a resampling analysis with 500 repetitions and varying *p* with varying means and standard deviations, $k_{sig.FFC}$ and $k_{sig.binom}$ were compared. At maximum, both k_{sig} differed by 14, i.e. the FFC would indicate a significant decline or recovery with 14 occurrences more or less than the binomial distribution.

Comparing the thresholds for significant recoveries or declines (k_{sig}), the intervals which indicate no significant change (i.e. the difference between $k_{sig.inc}$ and $k_{sig.dec}$) are generally narrower for FFC than the binomial distribution. This implies that k_{sig} determined by the binomial distribution is more precautionary when indicating recoveries, but less sensitive to declines than applying the FFC.

A RandomForest-model analysed the influence of the total number of samples (N), the mean and standard deviation of *p* (i.e. relative occurrence) on the deviation between $k_{sig.FFC}$ and $k_{sig.binom}$. The deviation between $k_{sig.FFC}$ and $k_{sig.binom}$ increases with mean and standard deviation of *p* (Fig. 5). However, the standard deviation of *p* has the most influence on the deviation between $k_{sig.FFC}$ and $k_{sig.binom}$, whereas the number of observations and mean of *p* have limited effect.

6.3. *Comparison with assessment methods*

The analysis comparison presented in Section 5.2 between changes in SSB and BOA indicates a high potential for agreement between both methods, with fewer disagreements observed than expected by chance. However, our results indicate that assessment outcomes of BOA can diverge from, and even contradict changes in SSB on occasion. Hence, species assessed by BOA might indicate a different status than when assessed by an analytical stock assessment. However, it is inevitable that some comparisons will fail as time series of occurrence and SSB can substantially differ for ecological reasons. For example, the frequency of occurrence of a highly abundant species can reach a plateau (at a maximum of 1) even though SSB might still vary substantially between periods. A diverging outcome may also be valid, e.g. where a species range has contracted but there is no change in biomass, and in such cases both methods/outcomes can be complimentary. Additionally, analytical stock assessments usually include landings or catch data, and so the different underlying data sources may also contribute to the differences in outcomes from the two assessment methods.

The advantage of the BOA lies in its generic applicability to presence-absence data and is applicable to situations when data are too sparse for other assessment methods, such as time series based assessments of abundance (Probst and Stelzenmüller, 2015), production models (Pedersen and Berg, 2017) or analytical stock assessments. We therefore suggest that the assessment of a species population/stock should be made using the most suitable statistical method given the data available.

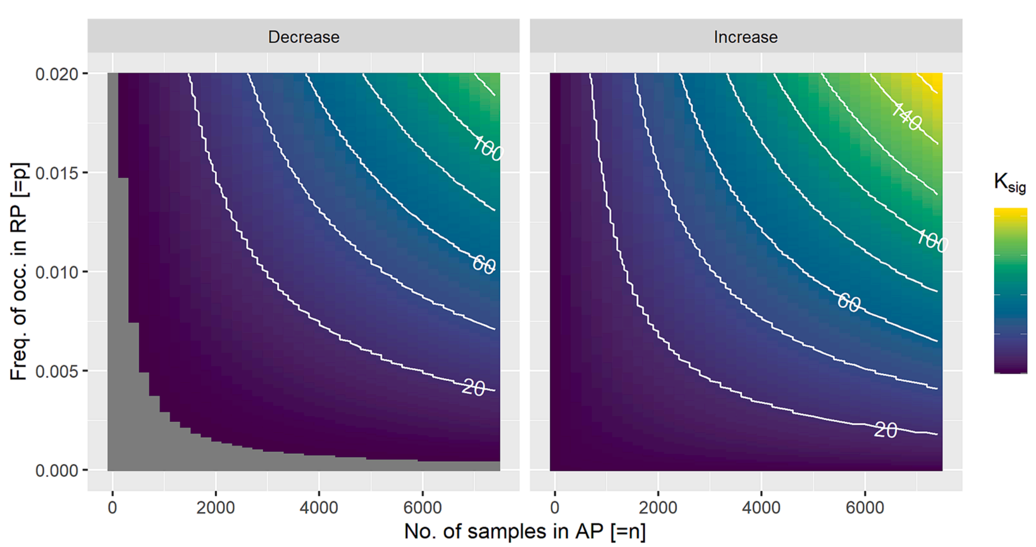

Fig. 4. The relationship between the number of hauls in the assessment period (AP), the frequency of occurrence (*p*) in the reference period (RP) and the significance limits for observed occurrences in the assessment period (K_{sig}) for predicting decreases (right) and increases (left). Note the grey area in the left panel indicating the inability to identify significant declines because the number of hauls and/or *p* is too low.

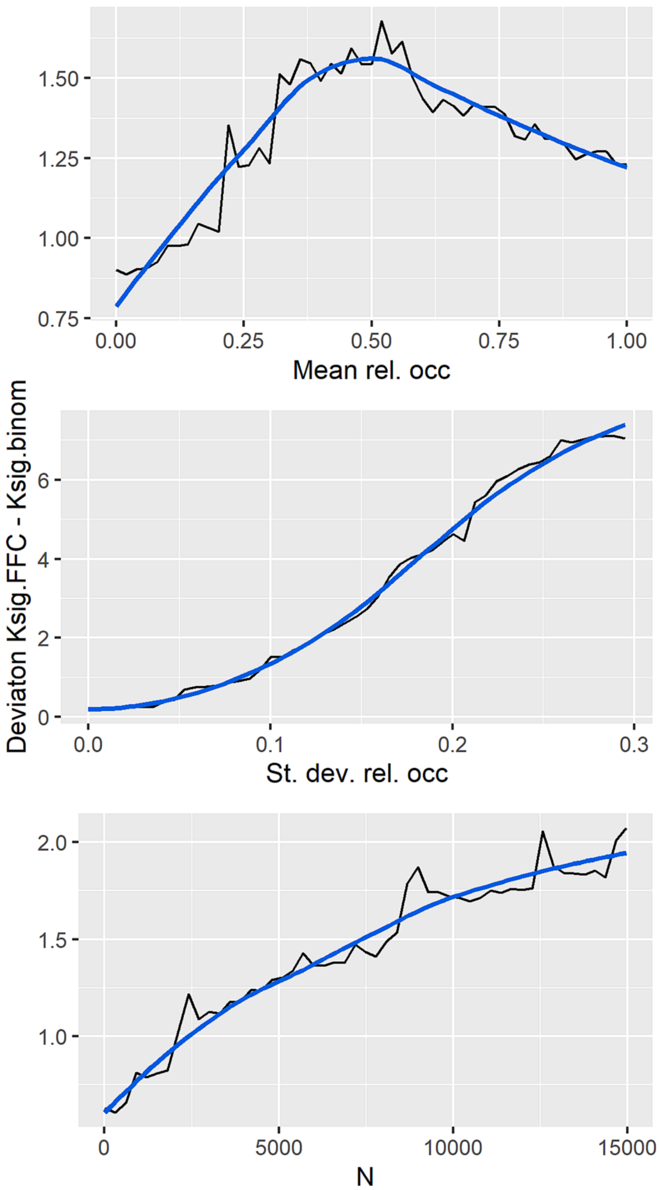

Fig. 5. Partial dependencies of a Random Forest-model analysis in the deviation between k_{sig} obtained from Fast Fourier Transformation (FFC) and the binomial distribution.

Accordingly, BOA is suitable for data-poor populations. For the OSPAR QSR 2023 assessment of the FC-1 indicator "Recovery of sensitive fish species", populations of sensitive species were excluded from the BOA where analytical stock assessments were available.

Further research comparing different assessment methods will be of interest for populations of fish, bird, mammal in the marine and terrestrial environments. For fish, the comparison against changes in SSB data from a greater suite of commercial species against BOA would be useful. Furthermore, the comparison of BOA against IUCN red list criteria such as changes in population abundance (Dulvy et al., 2021) to understand the comparability of BOA to other data-limited assessment approaches. Such analysis, however, was beyond the scope of this study as the alignment of temporal and spatial survey and assessment scales requires careful consideration and preparation.

6.4. *Choice of integration/selection method*

The choice of how to integrate or select the assessment information from all available survey time series can affect the overall conclusion on the status of a stock. Therefore, the integration and selection methods proposed here may require further development and validation to ensure an optimal integration of all available information. For the time being, BS seems to yield more consistent results in comparison to SSB than either BI or the ISCR (see Fig. 3). BI has the disadvantage that it can simultaneously indicate significant recoveries and declines (e.g. for turbot *Scophthalmus maximus* of shad species *Alosa spp* in RPL in the Greater North Sea case study) resulting in a "mixed" assessment outcome. These cases could be further resolved by including expert judgment or by adapting a precautionary approach that weighs the significant indication for declines more heavily. In contrast the I-Score and BS do not provide ambiguous outcomes and based on this feature might be preferred in future applications.

7. Conclusions

The major advantage of BOA is its generic applicability to time series of occurrence for rare or rarely observed species. Hence, it can be considered as a new assessment tool for the implementation of ecosystem-based management approaches that complements already existing assessment schemes. The application of BOA closes previous assessments gaps for rarer species, thereby providing an early warning signal for potentially declining populations that may otherwise have gone undetected and provide evidence for where further monitoring or management considerations are required. But care has to be taken when contradicting information from other assessment sources is available. BOA might be used if no other information on the status of a species is available, however, when abundance time series from stock assessments or other sources are available, results from BOA may provide complementary information to consolidate and validate primary assessment outcomes.

CRediT authorship contribution statement

W. Nikolaus Probst: Conceptualization, Methodology, Software, Formal analysis, Visualization, Writing – original draft, Funding acquisition. **Christopher P. Lynam:** Conceptualization, Methodology, Data curation, Writing – original draft, Funding acquisition. **Joanna K. Bluemel:** Methodology, Validation, Visualization, Writing – review & editing. **Maurice Clarke:** Supervision, Validation, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data are available online already and pointed to within the manuscript

Acknowledgements

The authors thank the OSPAR technical subgroup on fish for constructive comments on the development of the methodology for the assessment of sensitive fish. We also acknowledge the fruitful discussions at the ICES Workshop on the production of abundance estimates for sensitive species (WKABSENS) and the previous work by the ICES Workshop on Fish of Conservation and Bycatch Relevance (WKCOFIBYC). We also benefited by critical comments made by Ewen Bell (Cefas) on a draft of the manuscript.

Funding

CPL and JB were supported by the UK Government Department for Environment, Food and Rural Affairs (DEFRA). CPL was also supported by the Horizon Europe program: GES4SEAS ("Achieving Good Environmental Status for maintaining ecosystem services, by assessing integrated impacts of cumulative pressures"; grant agreement no. 101059877; www.ges4seas.eu). Through the contribution by WNP this study is part of the Thünen research initiative on the implementation of the Marine Strategy Framework directive (MSFD) "Definition and validation of indicators and initial assessment".

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecolind.2023.111084.

References

Arkema, K.K., Abramson, S.C., Dewsbury, B.M., 2006. Marine ecosystem-based management: from characterization to implementation. Front. Ecol. Environ. 4, 525–532.

Åström, S., Åström, J., 2022. Bumblebees and butterflies in Norway. Version 1.4. N. I. f. N. Research.

Borja, A., Elliott, M., Andersen, J.H., Berg, T., Carstensen, J., Halpern, B.S., Heiskanen, A.-S., et al., 2016. Overview of integrative assessment of marine systems: The ecosystem approach in practice. Frontiers in Marine. Science 3.

Curtin, R., Prellezo, R., 2010. Understanding marine ecosystem based management: A literature review. Mar. Policy 34, 821–830.

Dulvy, N.K., Pacoureau, N., Rigby, C.L., Pollom, R.A., Jabado, R.W., Ebert, D.A., Finucci, B., et al., 2021. Overfishing drives over one-third of all sharks and rays toward a global extinction crisis. Curr. Biol. 31, 4773–4787.

Edgar, G.J., Bates, A.E., Bird, T.J., Jones, A.H., Kininmonth, S., Stuart-Smith, R.D., Webb, T.J., 2016. New Approaches to Marine Conservation Through the Scaling Up of Ecological Data. Ann. Rev. Mar. Sci. 8, 435–461.

Gascuel, D., Coll, M., Fox, C., Guénette, S., Guitton, J., Kenny, A., Knittweis, L., et al., 2014. Fishing impact and environmental status in European seas: a diagnosis from stock assessments and ecosystem indicators. Fish Fish. 17, 31–55.

Greenstreet, S.P.R., Rogers, S.I., Rice, J.C., Piet, G.J., Guirey, E.J., Fraser, H.M., Fryer, R.J., 2011. Development of the EcoQO for the North Sea fish community. ICES J. Mar. Sci. 68, 1–11.

Greenstreet, S.P.R., Rossberg, A.G., Fox, C., Le Quesne, W.J.F., Blasdale, P., Boulcott, P., Mitchell, I., et al., 2012. Demersal fish biodiversity: species-level indicators and trends-based targets for the marine strategy framework directive. ICES J. Mar. Sci. 69, 1789–1801.

Gröger, J.P., Missong, M., Rountree, R.A., 2011. Analyses of interventions and structural breaks in marine and fisheries time series: Detection of shifts using iterative methods. Ecol. Ind. 11, 1084–1092.

Grumbine, R.E., 1994. What is ecosystem management? Conserv. Biol. 8, 27–38.

Halpern, B.S., Walbridge, S., Selkoe, K.A., Kappel, C.V., Micheli, F., D'Agrosa, C., Bruno, J.F., et al., 2008. A global map of human impact on marine ecosystems. Science 319, 948–952.

Halpern, B.S., Longo, C., Hardy, D., McLeod, K.L., Samhouri, J.F., Katona, S.K., Kleisner, K., et al., 2012. An index to assess the health and benefits of the global ocean. Nature 488, 615–620.

Harvey, C.J., Keble, C.R., Schwing, F.B., 2017. Implementing "the IEA": using integrated ecosystem assessment frameworks, programs, and applications in support of operationalizing ecosystem-based management. ICES J. Mar. Sci. 74, 398–405.

Helcom, 2013. HELCOM core indicators - Final report of the HELCOM CORSET project. Balt. Sea Environ. Proc. 136, 74.

ICES, 2021. Workshop on the production of abundancev estimates for sensitive species (WKABSENS). ICES Sci. Rep. 3, 96–128.

ICES, 2022. ICES Advice 2022. ICES Advice Publications, Collection https://doi.org/10.17895/ices.pub.c.5796935.

ICES.2022b. Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak (WGNSSK). ICES Scientific Reports. 4:43. 1321 pp. http://doi.org/10.17895/ices.pub.19786285.

Katsanevakis, S., Stelzenmüller, V., South, A., Sørensen, T.K., Jones, P.J.S., Kerr, S., Badalamenti, F., et al., 2011. Ecosystem-based marine spatial management: Review of concepts, policies, tools, and critical issues. Ocean Coast. Manag. 54, 807–820.

Korpinen, S., Laamanen, L., Bergstrom, L., Nurmi, M., Andersen, J.H., Haapaniemi, J., Harvey, E.T., et al., 2021. Combined effects of human pressures on Europe's marine ecosystems. Ambio 50, 1325–1336.

Levin, P.S., Fogarty, M.J., Murawski, S.A., Fluharty, D., 2009. Integrated ecosystem assessments: developing the scientific basis for ecosystem-based management of the ocean. PLoS Biol. 7, 23–28.

Lindegren, M., Dakos, V., Gröger, J., Gardmark, A., Kornilovs, G., Otto, S.A., Möllmann, C., 2012. Early detection of ecosystem regime shifts: A multiple method evaluation for management application. PLoS One1 7, 1–9.

Link, J.S., Browman, H.I., 2014. Integrating what? Levels of marine ecosystem-based assessment and management. ICES J. Mar. Sci. 71, 1170–1173.

Link, J.S., Nye, J.A., Hare, J.A., 2011. Guidelines for incorporating fish distribution shifts into a fisheries management context. Fish Fish. 12, 461–469. https://doi.org/10.1111/j.1467-2979.2010.00398.x.

Lynam, Ribeiro, 2022a. Occurrences of sensitive fish species in scientific trawl surveys of the Northeast Atlantic 1983-2020. Cefas, UK. V1. https://doi.org/10.14466/CefasDataHub.128.

Lynam, Ribeiro, 2022b. A data product derived from Northeast Atlantic groundfish data from scientific trawl surveys 1983-2020. Cefas, UK. V1. https://doi.org/10.14466/CefasDataHub.126.

Moriarty, M., Greenstreet, S.P.R., Rasmussen, J., 2017. Derivation of groundfish survey monitoring and assessment data product for the Northeast Atlantic Area. Scottish Mar. Freshw. Sci. 8, 240. https://doi.org/10.7489/1984-1.

OSPAR, 2017. Intermediate Assessment 2017.

Pedersen, M.W., Berg, C.W., 2017. A stochastic surplus production model in continuous time. Fish Fish. 18, 226–243.

Probst, W.N., Stelzenmüller, V., 2015. A benchmarking and assessment framework to operationalise ecological indicator based on time series analysis. Ecol. Ind. 55, 94–106.

R Core Team, 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Stelzenmüller, V., Letschert, J., Gimpel, A., Kraan, C., Probst, W.N., Degraer, S., Döring, R., 2022. From plate to plug: The impact of offshore renewables on European fisheries and the role of marine spatial planning. Renew. Sustain. Energy Rev. 158.

Thorson, J.T., Barnett, L.A.K., Anderson, E., 2017. Comparing estimates of abundance trends and distribution shifts using single- and multispecies models of fishes and biogenic habitat. ICES J. Mar. Sci. 74, 1311–1321.