ICES Journal of
Marine Science

ICES Journal of Marine Science (2021), https://doi.org/10.1093/icesjms/fsab242

Downloaded from https://academic.oup.com/icesjms/advance-article/doi/10.1093/icesjms/fsab242/6453069 by Fiskeridirektoratet. Biblioteket. user on 07 January 2022

Original Article

A simulation approach to assessing bias in a fisheries self-sampling programme

Thomas L Clegg, Edvin Fuglebakk, Kotaro Ono, Jon Helge Vølstad, and Kjell Nedreaas

Institute of Marine Research/Havforskningsinstituttet , Postboks 1870 Nordnes, NO-5817 Bergen, Norway

**Corresponding author: tel: 55 23 85 00; e-mail: tom.clegg@hi.no*

Clegg, T. L., Fuglebakk, E., Ono, K., Vølstad, J. H., and Nedreaas, K. A simulation approach to assessing bias in a fisheries self-sampling programme. – ICES Journal of Marine Science, 0: 1–12.

Received 5 October 2021; revised 15 November 2021; accepted 16 November 2021.

The hierarchical structure and non-probabilistic sampling in fisher self-sampling programmes makes it difficult to evaluate biases in total catch estimates. While so, it is possible to evaluate bias in the reported component of catches, which can then be used to infer likely bias in total catches. We assessed bias in the reported component of catches for 18 species in the Barents Sea trawl and longline fisheries by simulating 2000 realizations of the Norwegian Reference Fleet sampling programme using the mandatory catch reporting system, then for each realization we estimated fleet-wide catches using simple design-based estimators and quantified bias. We then inserted variations (e.g. simple random and systematic sampling) at different levels of the sampling design (sampling frame, vessel, and operation) to identify important factors and trends affecting bias in reported catches. We found that whilst current sampling procedures for fishing operations were not biased, non-probabilistic vessel sampling resulted in bias for some species. However, we concluded this was typically within the bounds of expected variation from probabilistic sampling. Our results highlight the risk of applying these simple estimators to all species. We recommend that future estimates of total catches consider alternative estimators and more conservative estimates of uncertainty where necessary.

Keywords: design-based, hierarchical sampling, random forest, reference fleet, self-sampling.

Introduction

Self-sampling by fishers is emerging as an effective method of collecting data at sea (Starr, 2010; Lordan *et al.*, 2011; Roman *et al.*, 2011; Kraan *et al.*, 2013; Bell *et al.*, 2017). An extension of this approach is a reference fleet, defined as a group of active fishing vessels with an enhanced data collection role requiring training and support (Mangi *et al.*, 2013). Reference fleets can reduce the logistics and costs of data collection compared with observer programmes (Mangi *et al.*, 2013; Suuronen and Gilman, 2020), and improve relationships with the fishing industry through participatory research and a two-way communication channel (Kraan *et al.*, 2013).

As a relatively new approach, data collection through self-sampling has received more scrutiny than other more established methods such as scientific observers (Kraan *et al.*, 2013; Mangi *et al.*, 2013) towards the representativeness of samples for making inferences about the wider fleets it covers. Multi-stage sampling of catches in fisheries results in complex, hierarchical data. At each stage, there are chances that bias is introduced in either a self-sampling or independent observer programme (ICES, 2008; Table 1), reducing the representativeness of samples.

The bias of an estimate is the degree to which the expected value, obtained through repeated sampling and estimation, differs from the true value (Jessen, 1978). The reality of fisheries sampling makes it difficult to assess the bias of total catch estimates as the true value is not known for comparison. Under a landing obligation, such as those implemented in Norway and the EU, reported catches do not always accurately reflect total extractions from the fishery. Catches may go unreported due to illegal discarding, intentional misreporting, or because of low resolution of reporting for some species or species identification errors (Pitcher *et al.*, 2002). However, no single study can address all aspects of bias. For example, comparing

© The Author(s) 2021. Published by Oxford University Press on behalf of International Council for the Exploration of the Sea. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Table 1. Potential sources of bias in a fishery sampling programme.

Aspect of sampling design	Potential sources of bias
Sampling frame	• A poorly defined sampling frame will affect representativeness of samples. • Insufficient data to define the sampling frame risks out-of-frame samples being included.
Sampling units	
1) Vessel	• Opt-in participation may bias towards more compliant vessels (rejection effect). • Type and size of compensation may influence motivation to participate. • Mandatory participation may influence fishing strategy if 100% of fishing activity is not observed (Liggins *et al.*, 1997; Benoît and Allard, 2009; Snyder and Erbaugh, 2020).
2) Fishing operation	• Human selection of sampled fishing operations may favour convenience (e.g. sampling smaller hauls to reduce time spent sampling) and introduce bias.
3) Catches	• Gear characteristics, variable habitat, and fishing strategy will likely result in a non-random distribution of fish throughout the catch operation, which if sampled opportunistically would produce biased observations (e.g. sampling from the first available portion of the catch). • Intentional manipulation of catch data, possibly where results could have a large impact on management or policy decisions (e.g. misreporting catches of protected species or species with limited quota). • Poor sampling techniques, either through inadequate training (e.g. species identification, misuse of equipment) or lack of time during catch processing.
Estimator	• Choice of estimator must be upheld by the relevant assumptions

self-sampling data with independent observations of known reliability (Faunce, 2011; Roman *et al.*, 2011) will only address measurement error such as under-reporting of sensitive species. Similarly, comparing multiple estimators or sampling designs may address biases in those specific aspects (Diamond, 2003; Cahalan *et al.*, 2015; Cahalan and Faunce, 2020), but prior knowledge of potential biases will help when defining the candidate estimators.

Total catches can be broken down into the portion reported to the authorities (in daily logbooks or landing reports), and the portion that is not reported but does occur. The unreported component is especially problematic because it is difficult, if not impossible to quantify biases related to them (Ainsworth and Pitcher, 2005). However, we can reach a better understanding of biases affecting estimates of total catches by focusing on just the reported component. Mandatory catch reporting acts as a census of fishing effort and the reported component of total catches, so it can be utilized to explore such biases, assuming that biases affecting reported catches are also likely to affect total catches (Liggins *et al.*, 1997).

The Institute of Marine Research recruits vessels and maintains the Norwegian Reference Fleet, a fisheries self-sampling project, which tasks participating vessels with regularly gathering of data on fishing operations during normal fishing activity (Clegg and Williams, 2020). An independent evaluation of the Norwegian Reference Fleet by Bowering *et al.* (2011) concluded that based on expert judgement and limited analyses, the sampling programme is representative of the wider fleets it covers. However, Bowering *et al.* (2011) concluded that focused analyses are needed to evaluate representativeness of individual segments of the Norwegian Reference Fleet.

This study aims to understand the representativeness of the Norwegian Reference Fleet sampling design by identifying biases that are likely to affect estimates of total catches. We addressed this by focusing on the reported component of total catches, for which we have census data in the form of mandatory daily catch logbooks. We simulated Norwegian Reference Fleet sampling design in the Barents Sea trawl and longline fisheries between 2012 and 2018. We then estimated fleet-wide catches based on the reports in the sample and quantified bias through comparison with observed reported catches and used random forest models to understand which variables are important for explaining variations in bias. Assuming biases in our estimates of reported catches will likely affect estimates of total catches, we discuss the results in the context of total catches to suggest ways in which the biases can be reduced or mitigated.

Data

Case study fisheries

Our case study is focused on a portion of the Barents Sea bottom trawl and longline fisheries, defined as vessels with overall length (LOA) greater than 28 m using bottom trawl or longline fishing gears to target demersal fish species in the statistical areas highlighted in Figure 1. Area 24 only includes the trawl fishery, as the longline fishery does not extend into this statistical area. In the trawl fishery, a fishing operation is defined as a haul, whilst in the longline fishery it is defined as all hooks hauled from all longlines in a calendar day.

Both fisheries occur all year round, peaking between November and January. However, statistical areas 23 and 24 are typically inaccessible between January and April due to sea ice. Vessels predominantly target cod (*Gadus morhua*) and to a lesser extent haddock (*Melanogrammus aeglefinus*), but also infrequently target saithe (*Pollachius virens*), tusk (*Brosme brosme*), Greenland halibut (*Reinhardtius hippoglossoides*), and beaked redfish (*Sebastes mentella*).

The northern prawn (*Pandalus borealis*) trawl fishery and pelagic trawl fisheries (mainly capelin, *Mallotus villosus*) carry a small risk of being included in this study because of overlap in space and time and possible erroneous reporting of trawl gear codes. However, these fisheries are easily identifiable due to very high selectivity, allowing fishing operations to be removed if the dominant species was not a demersal fish.

For each year, we post-stratified samples based on a combination of statistical area (Figure 1) and season (winter: January–April; summer: May–August; and autumn: September–December). Three seasons reflect the seasonality of the Barents Sea, such as ice cover restricting access to areas 23 and 24 (Figure 1) in winter. Furthermore, there were insufficient data to estimate fleet-wide reported catches on a monthly timescale. We only estimated reported catches in strata with three or more fishing operations sampled, as unsam-

Downloaded from https://academic.oup.com/icesjms/advance-article/doi/10.1093/icesjms/fsab242/6453069 by Fiskeridirektoratet. Biblioteket. user on 07 January 2022

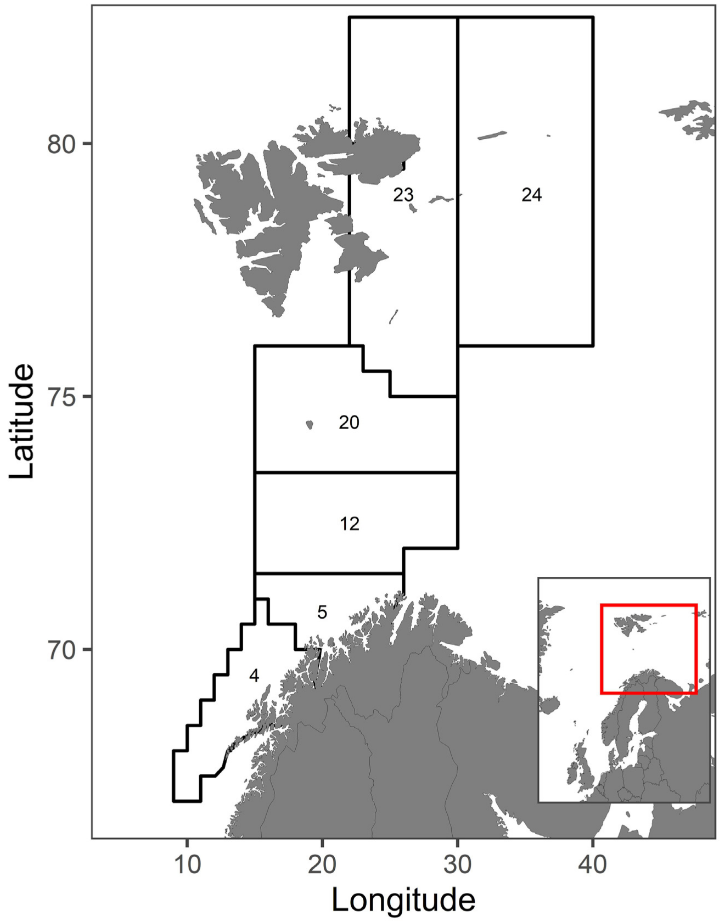

Figure 1. Statistical areas in the Barents Sea trawl and longline fisheries as defined by the Norwegian Directorate of Fisheries. Area 24 is excluded from the longline fishery due to negligible fishing activity.

pled strata require imputation methods that are often subjective and so introduce new biases that are difficult to quantify (Stratoudakis *et al.*, 1999; Lohr, 2010).

Table 2 provides a full list of species included in this study. Broad species groups were removed from the study (e.g. unidentified flatfish) to avoid double counting. An exception to this rule was skates and rays for which species identification is notoriously difficult and are typically grouped in the reported catches. Species observed in less than 1% of fishing operations in any given year were removed from the study. Extremely rare species typically require more complex modelling approaches for estimating total catches, so the relevance of bias using design-based estimators is not of relevance to this study.

Norwegian reference fleet

The Norwegian Reference Fleet project is a trust-based collaboration between fishers and scientists to improve data for input into stock assessments and provide data on bycatches and discards. Our selected fisheries in this study are prioritized for the Norwegian Reference Fleet, meaning that active participation is required in the contract to ensure representativeness and sufficient data for stock assessments.

Table 2. Species included in this study. Asterisks mark species included only in the longline fishery.

Common name	Scientific name	FAO code
Atlantic wolffish	*Anarhichas lupus*	CAA
Northern wolffish*	*Anarhichas denticulatus*	CAB
Spotted wolffish	*Anarhichas minor*	CAS
Ratfish*	*Chimaera monstrosa*	CMO
Atlantic cod	*Gadus morhua*	COD
Greater forkbeard*	*Phycis blennoides*	GFB
Greenland halibut	*Reinhardtius hippoglossoides*	GHL
Haddock	*Melanogrammus aeglefinus*	HAD
Atlantic halibut	*Hippoglossus hippoglossus*	HAL
Common ling	*Molva molva*	LIN
Monkfish*	*Lophius piscatorius*	MON
Saithe	*Pollachius virens*	POK
Skates and rays*	*Rajidae*	RAJ
Beaked redfish	*Sebastes mentella*	REB
Golden redfish	*Sebastes norvegicus*	REG
Roughhead grenadier*	*Macrourus berglax*	RHG
Lesser redfish*	*Sebastes viviparus*	SFV
Tusk	*Brosme brosme*	USK

Downloaded from https://academic.oup.com/icesjms/advance-article/doi/10.1093/icesjms/fsab242/6453069 by Fiskeridirektoratet. Biblioteket. user on 07 January 2022

Vessel owners can apply to participate in the Norwegian Reference Fleet through a publicly open tender process, as mandated by law. This process involves a public announcement of a paid contract for a vessel to sample catches from their normal fishing activity over a 4-year period. Each tender specifies a list of mandatory and desired requirements, which aim to recruit a typical vessel in the defined category. Applicants must provide evidence that they meet these requirements to be included in the selection process. Eligible applications are then assessed by a panel to evaluate the desired requirements. If after this evaluation there are multiple eligible vessels, then the contract is awarded randomly.

The vessel categories relevant to this study are the trawl and longline categories where the tender requires that vessels must be greater than 39 m and 35 m LOA, respectively, and hold permit and quota to fish various demersal species in the Barents Sea fisheries. Over the study period, around 14% of all vessels in the study fisheries have participated in the Norwegian Reference Fleet (longline: 8/55 vessels; trawl: 6/42 vessels).

Each vessel has designated crew who are given training on the sampling protocol and species identification. These crew sample total catches systematically, such that one fishing operation is sampled every 2 days (1-in-2 systematic sample; Lohr, 2010). The starting day for systematic sampling is selected randomly for each trip by the crew or skipper. For trawl vessels, the fishing operation is defined as a single haul. A haul is randomly selected from all those planned on the sampling day. On longline vessels, for which fishing operation is defined as a calendar day, it is relatively easier to subsample the catch in a fishing operation. All specimens are recorded from a subsample of consecutive hooks, spanning the start, middle, and end of longlines that the crew or skipper deems representative of the catch composition for that day.

Daily logbooks

Norwegian law requires all vessels in the Barents Sea fisheries at or above 15 m LOA to report the weight of each species caught in every fishing operation through an electronic reporting system (ERS). Each entry contains an estimated total live weight for each species, alongside the time and location of the fishing operation. This at-sea reporting of total catches may be biased downwards if fish are discarded during processing or processed catches are misreported (i.e. illegally landed). Official catch statistics are reported as round weight (live weight when removed from the sea), but on factory vessels, all catch reporting is done post-production. Therefore, product weights are converted back to round weight using official conversion factors for each product (Norwegian Directorate of Fisheries, 2021). This weight conversion has negligible impact on our study as factors are applied consistently across all vessels. Although reported weights are estimated at sea, various regulations ensure that weights match those officially declared in sales notes when weighed and sold on land (10% tolerance) to verify the accuracy of reported catches (Gezelius, 2006; Gullestad *et al.*, 2015). Whilst catches are weighed more accurately in sales notes, they offer insufficient data resolution for our analysis because they are a summary of all fishing activity in each trip, which spans multiple statistical areas over a period of weeks. We, therefore, can view the ERS data as a census of true reported catches in the fisheries for the purposes of this study.

We extracted ERS logbook data, provided by the Norwegian Directorate of Fisheries, between 2012 and 2018 from all commercial fishing operations (longline: 21807; trawl: 109801) in the two case study fisheries. We removed 17 entries from trawl vessels for which trawl duration could not be calculated because erroneous start and stop times were reported. We also removed one anomalous entry where a trawl vessel reported 40 tonnes of fish for 1 min of fishing time.

Methods

In order to identify biases that are likely to affect estimates of unreported catches in the Barents Sea trawl and longline fisheries between 2012 and 2018, we performed a simulation study using the ERS logbook data on reported catches. By simulating the Norwegian Reference Fleet sampling design using data on reported catches, we can generate estimates of the reported component of catches that can be directly compared to the true values, such that biases can be quantified. Results from these simulations can then be applied to improve estimates of total catches (i.e. reported + unreported) using the real data generated by the Norwegian Reference Fleet sampling programme mimicking the Norwegian Reference Fleet sampling design and various other designs (e.g. simple random), and quantified bias through comparison of the estimated fleet-wide catches from the simulated sample with observed reported catches.

Simulating sampling designs

The simulation framework consists of three components which we manipulated: the sampling frame, vessel sampling, and fishing operation sampling. Simulating these in a fully-crossed design (Figure 2) ensured balanced groups for the statistical analysis (Boulesteix *et al.*, 2012).

We assessed if the Norwegian Reference Fleet is representative of the entire fishery (> 28 m LOA) by first estimating total reported catches of vessels within the length ranges defined in Norwegian Reference Fleet tender specifications (sampling frame: *LIM*; Figure 2) and compared it to an estimation of reported catches from all vessels in the study fishery (sampling frame: *ALL*).

To evaluate the bias incurred from the non-probabilistic selection of vessels in the current Norwegian Reference Fleet (vessel: *RF*; Figure 2), we also performed two probabilistic selections of vessels for comparison. First, a simple random sample (vessel: SRS_V) of vessels within the vessel length class requirements of the Norwegian Reference Fleet tender specifications, and second, a weighted random sample (vessel: *WRS*) where vessels were selected with a probability proportional to the fishing effort (fishing days) in the previous year.

Finally, we simulated the 1-in-2 systematic sampling protocols by the Norwegian Reference Fleet (fishing operation: *SYS*; Figure 2), then compared it to a simple random sample (fishing operation: SRS_{FO}) from all fishing operations by sampled vessels in each stratum, with a sample size equal to that of the post-stratified systematic sample for each simulation. Systematic sampling is expected to be the equivalent of simple random sampling but was nevertheless included for confirmation.

Estimation procedure

We chose two conventional design-based estimators [Equations (1) and (2)] that are currently used for estimating discards and bycatches in Norwegian fisheries (Bærum *et al.*, 2019; Berg and Nedreaas, 2020; Moan *et al.*, 2020). These simple estimators assume that samples are randomly selected from all fishing operations in

Downloaded from https://academic.oup.com/icesjms/advance-article/doi/10.1093/icesjms/fsab242/6453069 by Fiskeridirektoratet. Biblioteket. user on 07 January 2022

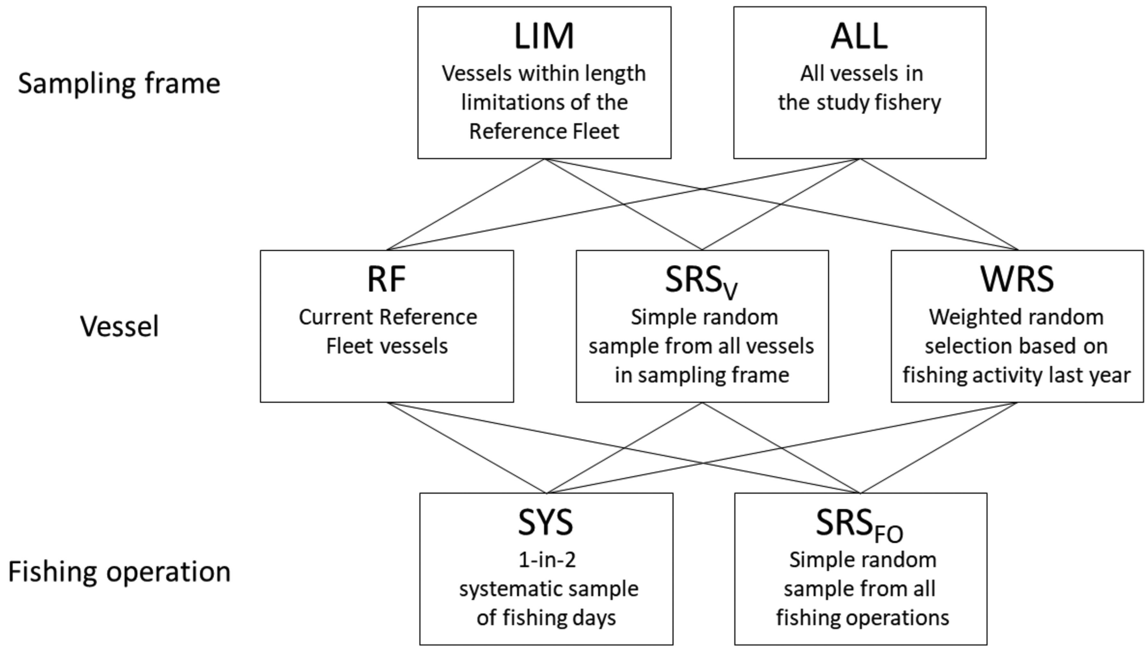

Figure 2. Schematic diagram of the simulations showing the three stages and fully-crossed design.

each stratum, effectively ignoring variations in catches between vessels. Despite the hierarchical sampling design of the Norwegian Reference Fleet, it is not yet clear how a multi-stage estimator should be defined due to the non-probabilistic selection of vessels and a lack of understanding of which levels of the sampling design contribute the most to variations in total catches. Therefore, for the purposes of simplicity, we ignored the hierarchical sampling design and applied commonly used simple estimators.

For species i in year j, total reported catches ($\hat{Y}_{i,\ j}$) in sampled strata were estimated using the stratified unit estimator [Equation (1)] for the longline fishery, and the stratified ratio estimator [Equation (2)] for the trawl fishery:

$$\hat{Y}_{i,j} = \sum_{k=1}^{K} \frac{N_{j,k}}{n_{j,k}} y_{i,j,k}, \tag{1}$$

$$\hat{Y}_{i,j} = \sum_{k=1}^{K} \frac{y_{i,j,k}}{x_{j,k}} X_{j,k}, \tag{2}$$

where $k =$ stratum, $y =$ weight of reported catches sampled, $n =$ number of fishing operations sampled, $N =$ total number of fishing operations in population, $x =$ sampled trawl duration, and $X =$ total trawl duration in population. Note that when simulating the Norwegian Reference Fleet sampling design for longline vessels, we do not need to account the subsampling of hooks (which is done for Norwegian Reference Fleet sampling), as catches for the entire fishing operations are reported in the ERS logbooks.

The accuracy of the ratio estimator [Equation (2)] is partially dependent on the correlation between catch weight (y) and trawl duration (x). We, therefore, calculated the Spearman's rank correlation coefficient between catch weight and trawl duration for each species (Rochet and Trenkel, 2005; Lohr, 2010).

The re-sampling and estimation processes were repeated 2000 times (S = 2000), and the relative error (RE) of $\hat{Y}_{i,\ j}$ calculated as a measure of bias using:

$$RE_{i,j} = \frac{1}{S} \sum_{s=1}^{S} \frac{\hat{Y}_{i,j,s} - Y_{i,j,s}}{Y_{i,j,s}}, \tag{3}$$

where $Y_{i,\ j,\ s}$ is the observed annual catches of species i in sampled strata in year j and in simulation s.

Systematic sampling of fishing operations for each vessel will result in a different sample size for every realization of the sampling protocol. As the starting day for systematic sampling is selected for each vessel per simulation, the number of fishing operations in each stratum will vary slightly depending on where the sampling days fall, which is in turn proportional to the number of vessels in that year. For trawl vessels, this is further affected by which haul is selected for sampling on each day. These variations were not controlled as they reflect the true variations that arise from sampling protocols.

Modelling the sources of bias

We used random forests (Breiman, 2001) as a regression method to explore which aspects of the sampling process are causing bias in estimations of reported catches. For a detailed explanation and guidance for random forest models in various biological and ecological contexts, we recommend Cutler *et al.* (2007), Boulesteix *et al.* (2012), Fox *et al.* (2017), and Siders *et al.* (2020).

First and foremost, we chose random forests for a unique and novel feature: variable importance (Genuer *et al.*, 2009). Fitting a random forest model involves repeatedly sampling observations and explanatory variables from the original dataset. Those observations not included in model fitting (known as "out-of-bag" samples) can be used to test the accuracy of model prediction (comparable to cross validation in Generalized linear models). However, random forests can also examine how the exclusion of each explanatory variable affected the accuracy of prediction. This overall decrease in accuracy can be used as a measure of importance for each

Downloaded from https://academic.oup.com/icesjms/advance-article/doi/10.1093/icesjms/fsab242/6453069 by Fiskeridirektoratet. Biblioteket. user on 07 January 2022

Table 3. Variables considered in the random forest models.

Variable	Type	Description
Vessel	Categorical	Vessel sampling design simulated†
Fishing operation	Categorical	Fishing operation sampling design simulated†
Sampling frame	Categorical	Population from which vessels are sampled†
Species	Categorical	Species or species group (Table 2)
Year	Categorical	Year (2012–2018)
Encounter rate	Continuous	Proportion of fishing operations where the species was observed (mean across all simulations).
Sample size*	Continuous	Number of fishing operations sampled annually (mean across all simulations)
Correlation coefficient*	Continuous	Spearman's rank correlation coefficient between catch weight and trawl duration for each species and year

†See Figure 2 for categories.
*Only used in trawl fishery model as it is only relevant to the bias of a ratio estimator.

explanatory variable. For example, if there was no change in model accuracy when a variable was not included, then that variable was not important in explaining any variations in bias.

Random forest methods also suit the exploratory nature of our study, mitigating against the bad practice of "data dredging" (Burnham and Anderson, 2002). There is no iterative variable selection and testing process, meaning all potential variables are included in a single model. Similarly, random forests automatically capture non-linear relationships and complex interactions between variables without the need for prior specification.

To account for possible biases arising from differences in scale and type of predictor variables, we used a class of decision trees called conditional inference trees to build the random forest (Hothorn *et al.*, 2006; Strobl *et al.*, 2007). All models were fitted using the R package *party* (version 1.3–7; Hothorn *et al.*, 2006; Strobl *et al.*, 2007, 2008) with "*cforest_unbiased*" convenience controls. Importance was estimated conditionally, which accounts for correlated variables and interactions (Strobl *et al.*, 2008, 2009). We also explored the relationship between bias and each explanatory variable using partial dependence plots (Friedman, 2001), calculated using the R package *edarf* (version 1.1.1; Jones and Linder, 2016).

In addition to the various elements of sampling design we simulated (Figure 2), we also included five other predictor variables in the random forest analysis: species, encounter rate, year, sample size, and correlation coefficient (Table 3). Sample size and correlation coefficient were only included in the trawl fishery model where we used a ratio estimator, for bias is a function of these two variables. As an alternative to encounter rate, we considered total annual reported catches of all species. However, encounter rate was preferred as a better indication of the rarity of species and was more relevant to the sampling data than the population data.

We fitted a random forest model to each fishery independently. Model tuning was steered by two hyperparameters: number of trees (*ntree*) and number of randomly sampled variables at each split (*mtry*), which were optimized using a simplified grid search to minimize the out-of-bag mean square error (MSE). We tested all possible *mtry* values (longline: 1–6; trawl: 1–8 variables), and five *ntree* values spaced evenly between 50 and 1000. Results may be sensitive to the random seed, so we compared variable importance outputs from five initial runs with new random seeds. Any substantial changes in results indicates instability, meaning more trees should be added. Tuning resulted in both models being fitted using $mtry = 4$ and $ntree = 500$ (Supplementary Table S1).

Additional analyses

The current selection of vessels in the Norwegian Reference Fleet is the only possible realization of the expert judgement selection process. Comparatively, we have simulated a large number of realizations using a simple random sample of vessels. We assume the non-probabilistic, expert judgement selection behaves like a simple random sample to be able to use design-based estimators. For this assumption to be upheld, we would expect that the estimate based on expert judgement selection of vessels would lie within the distribution of estimates based on a simple random selection. We evaluated this using z-scores for each estimate of relative error, taking the mean and standard deviation from the distribution of relative errors from simulations using a simple random sample of vessels. In addition to this, we made pairwise comparisons of accuracy using the different vessel sampling methodologies for individual simulations to determine how often a simple random sample out-performs the expert judgement selection.

Results

Amongst the three components of the catch sampling design (Figure 2), vessel sampling method was the most important contributor of bias when estimating the reported component of total catches in both the longline and trawl fisheries (Figure 3). With the estimators used here, the current realization of the Norwegian Reference Fleet vessel selection procedure tends to overestimate catches when averaged across all species (Figure 4a). However, this summary statistic does not account for the important variations in bias across years and species (Figure 4b and c).

Figure 5 illustrates to what extent the expert judgement selection in the current Norwegian Reference Fleet behaves like simple random sampling. More than half of the annual species catch estimates using a non-random vessel selection were within one standard deviation of the mean from a simple random sample (z-score < 1; longline: 56%; and trawl: 77%), with few estimates being outside the 95% confidence interval (z-score > 1.96; longline: 12%; and trawl: 4%). However, the distribution of z-scores was skewed to the right (Figure 5), confirming the tendency for this non-random sample of vessels to result in overestimation in reported catches of individual species.

As the sampling frame was not important in explaining the variations in bias when estimating the reported component of catches for individual species (Figure 3), it suggests that the vessel length

Downloaded from https://academic.oup.com/icesjms/advance-article/doi/10.1093/icesjms/fsab242/6453069 by Fiskeridirektoratet. Biblioteket. user on 07 January 2022

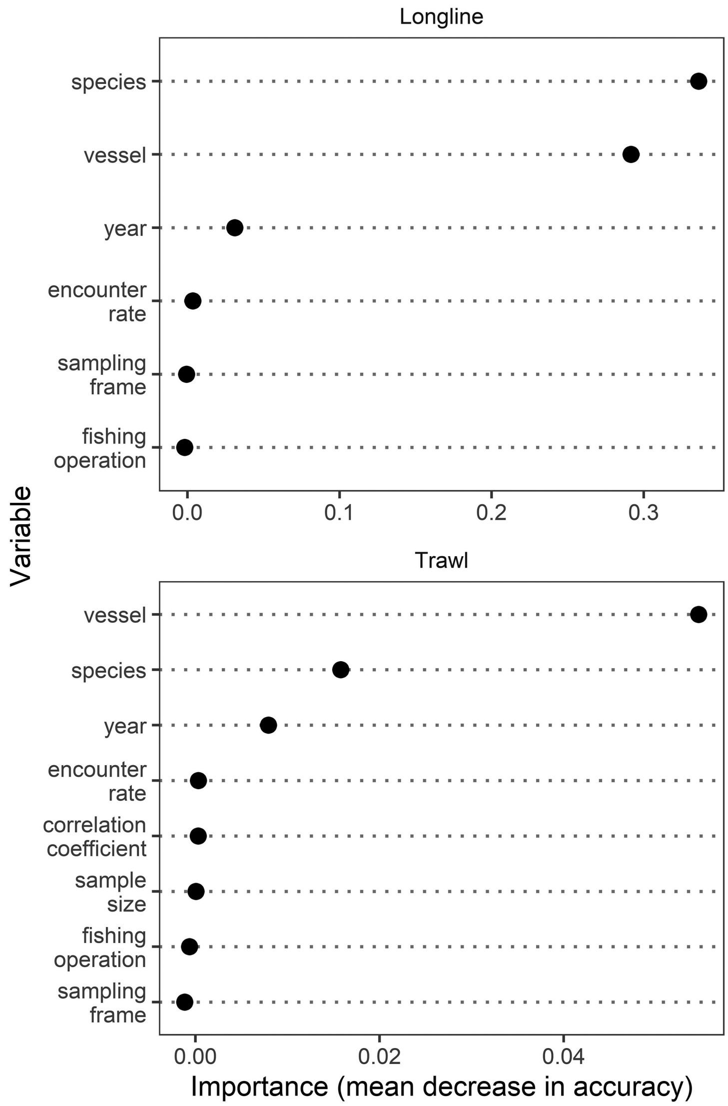

Figure 3. Importance of variables for explaining variations in relative error when predicting the reported component of total catches. Measures of conditional importance of variables estimated using random forest models. Note different scales on *x*-axes.

requirements in the Norwegian Reference Fleet do not affect the representativeness in relation to all vessels in the fishery (> 28 m LOA). Similarly, the simulated methodologies for sampling fishing operations were of negligible importance when estimating total reported catches (Figure 3), confirming that systematic sampling of fishing operations is the equivalent of simple random sampling.

Variations in relative error of estimated reported catches between species was of high importance in both fisheries and was the most important variable in the longline fishery (Figure 3). These variations across species were large in both fisheries but was most extreme in the longline fishery (Figure 4c). For many species, the probabilistic sample of vessels still resulted in estimation bias, which further highlights that the weak assumptions in our chosen estimators (i.e. not applying multi-stage estimators, and poor correlation between catches and trawl duration) resulted in bias due to species-specific factors. Total reported catches were underestimated only in a small number of cases, most notably beaked redfish (REB) and skates and rays (RAJ) in the longline fishery.

In the trawl fishery, biases were consistently largest across species when using the non-random selection of vessels (Figure 4c). The ratio estimator [Equation (2)] is known to be biased when the correlation between catches and trawl duration is low, and with a smaller sample size (Lohr, 2010). However, the random forest model indicated both these factors were of low importance when explaining variations in bias (Figure 3). Given the small variations in these explanatory variables (Supplementary Figure S1), this does not provide insights into how bias would improve with higher correlation coefficients or larger sample size.

Despite accounting for many other variables that could explain differences in relative error, there were still annual variations affecting bias of reported catch estimates across all species, albeit weak (Figure 3). These variations were more extreme for non-random Norwegian Reference Fleet vessel sampling (Figure 4b) than for probabilistic sampling of vessels.

Downloaded from https://academic.oup.com/icesjms/advance-article/doi/10.1093/icesjms/fsab242/6453069 by Fiskeridirektoratet. Biblioteket. user on 07 January 2022

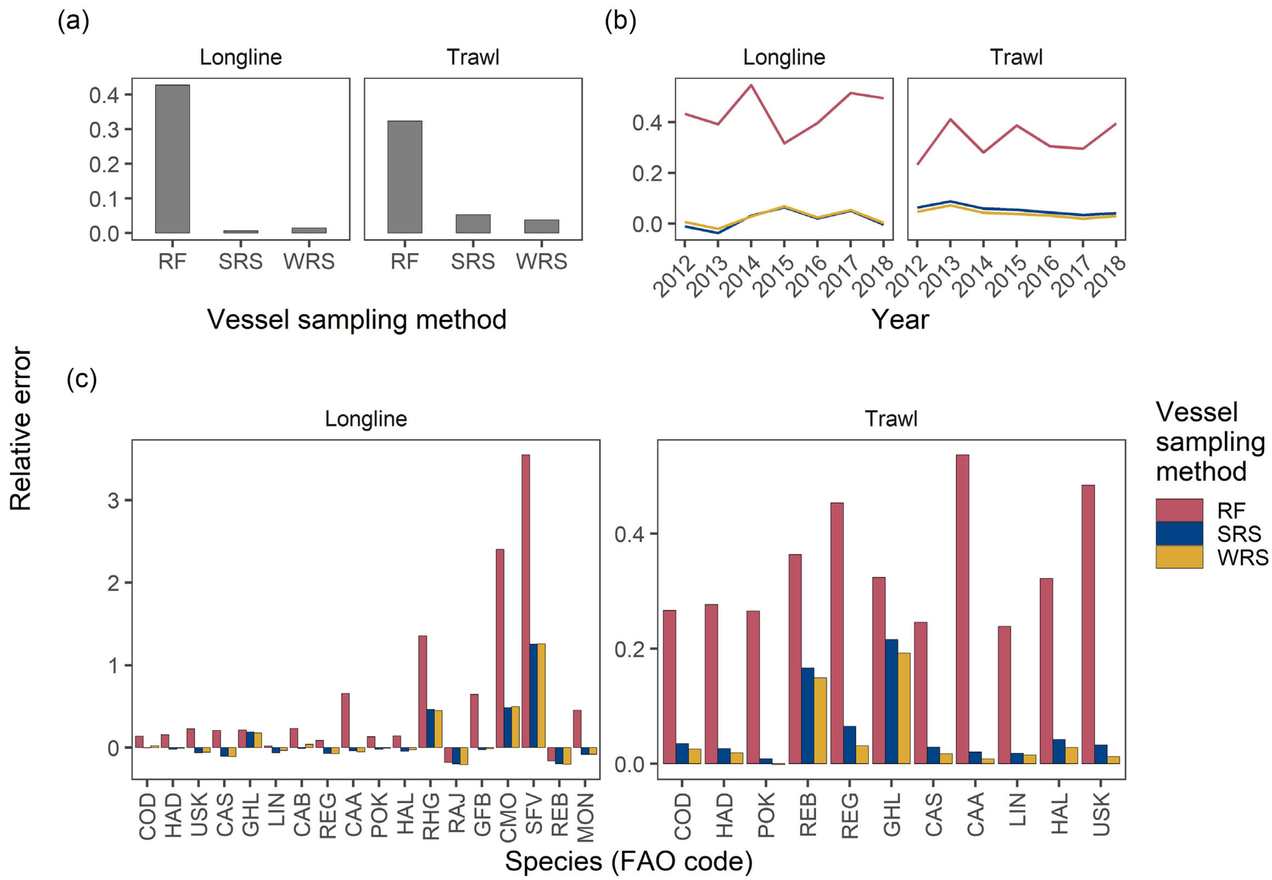

Figure 4. Partial dependence plots of the most important variables in the random forest models, showing the marginal effect of variables on relative error. (a) Vessel selection (RF = current Norwegian Reference Fleet vessels (fixed); SRS = simple random sample; and WRS: weighted random sample). (b) and (c) include interactions between the three vessel sampling methodologies. Note different scales on *y*-axes in panel c. Species ordered by descending total reported catches for each fishery. Species names are listed in Table 2.

Discussion

Through a simulation approach using the reported component of total catches, we have identified aspects of the Norwegian Reference Fleet vessel selection and catch sampling design that will likely affect biases in estimates of total catches for a range of species in the Barents Sea trawl and longline fisheries. We have identified that the vessel selection process for the Norwegian Reference Fleet resulted in an overestimation of the reported component of total catches for many species, indicating possible biases in similar estimations of unreported catches. However, we also found biases in reported catch estimates using probabilistic sampling, which indicates the hierarchical structure of Norwegian Reference Fleet programme's sampling design does not meet the assumptions of the simple estimators used here. Furthermore, important variations between species indicate that the suitability of our chosen estimators is highly dependent on species-specific factors, suggesting the consideration of alternative estimators.

Vessel selection process

There is potential for the vessel selection process to result in bias due to practical difficulties in maintaining a reliable at-sea sampling programme such as issues regarding participation (Vølstad and Fogarty, 2006; Benoît and Allard, 2009; Kraan *et al.*, 2013) and budget restrictions (Borges *et al.*, 2004). For a simple random selection of vessels, one realization could risk selecting vessels with very different fishing strategies relative to the wider fleet. Furthermore, randomly selected vessels may be rarely active in the fishery, yielding little data for a large investment in equipment and training. A weighted random sample could reduce those risks but could also introduce further complications. For example, a vessel that is consistently active in the fishery could spend long periods in the harbour for repairs or refurbishments, reducing its selection probability for the following year. Probabilistic sampling implies that all vessels are willing to participate if selected, which is unrealistic if not impossible, as current laws require that all vessels must have the opportunity to apply to an open tender, rather than being selected and requested to participate.. Mandatory participation can function with independent samplers (Ewell *et al.*, 2020), but self-sampling requires large amounts of time and effort alongside normal fishing activity. If self-sampling was mandatory and without compensation, then we would expect trust to deteriorate between scientists and fishers and a reduction in data quality (Jacobsen *et al.*, 2012; Mangi *et al.*, 2016). Our results found that for the majority of cases, an estimate using expert judgement selection of vessels for the Norwegian Reference Fleet was within the range of expected estimates using a simple random selection of vessels (Figure 5). However, there is a tendency when using Norwegian Reference Fleet to overestimate reported catches, which can be significantly large for certain species.

Downloaded from https://academic.oup.com/icesjms/advance-article/doi/10.1093/icesjms/fsab242/6453069 by Fiskeridirektoratet. Biblioteket. user on 07 January 2022

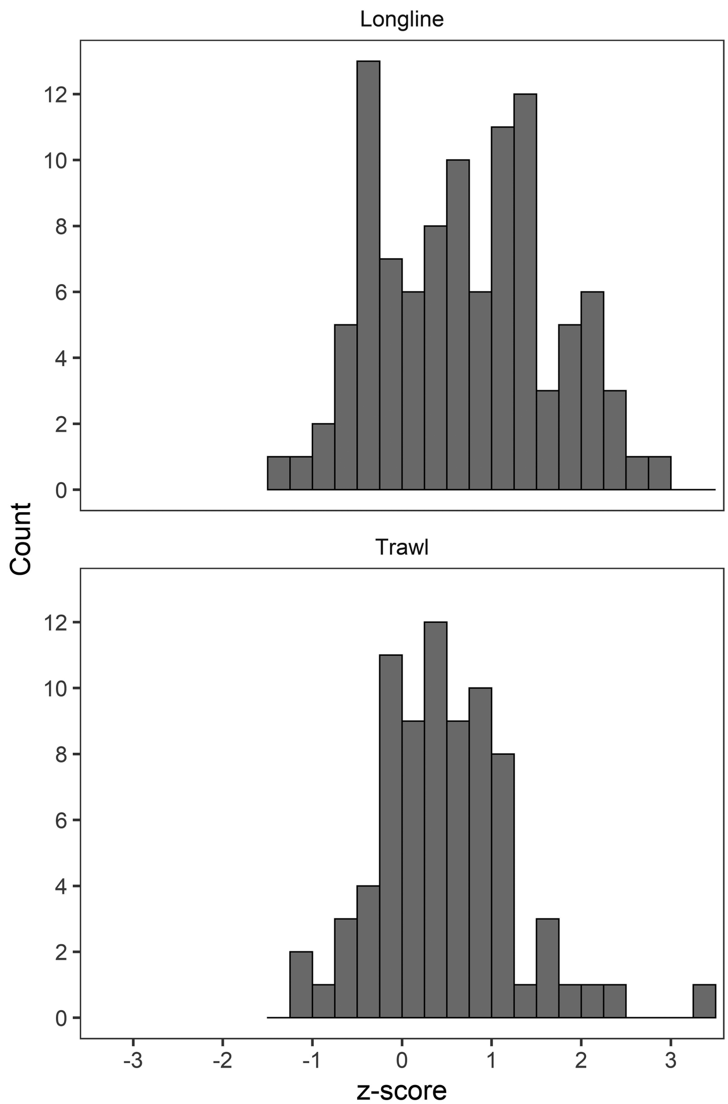

Figure 5. The range of z-scores for bias in annual estimates of reported catches for species using an expert judgement selection of vessels. A z-score is the number of standard deviations from the mean of simulated estimates using a simple random sample of vessels (i.e. where an expert judgement selection lies within a distribution of simple random samples). Each z-score represents an estimate of annual reported catches for one species.

Supplementary Figures S2 and S3 in the Supplementary materials provide comparisons of fishing activity between Norwegian Reference Fleet vessels and the wider fleet to suggest reasons for the tendency of Norwegian Reference Fleet sampling to overestimate catches. Norwegian Reference Fleet vessels have higher catch rates and different fishing strategies compared to the wider fleet. These differences could be caused by higher engine power and annual quotas, both of which will affect overall catch composition, and therefore, unreported catches. Furthermore, in selecting a small sample of vessels that are to be fixed for several years, the selection cannot be optimized to capture the tails of statistical distributions that may be derived from these vessels or their catches. It is reasonable to expect this to lead to some underestimation of variability of total catches, and some over- or under-estimation of total catches when the distribution of total catches is not symmetric.

Although a devoted study is required to provide evidence for the differences in catch rates, we can still consider ways to improve incentives for participating in the Norwegian Reference Fleet. A larger number of applications would improve the outcome of expert selection, and in the situation where similar vessels are short-listed, then random selection could be applied more effectively.

Variations in bias across species

We cannot discuss the biases in estimates for individual species in this exploratory study, but we can nevertheless discuss generalizations. Each species and stratum can be viewed as a domain (subpopulation), as these properties are unknown before the sample is taken (Lohr, 2010). Whilst there are methods to approximate the sampling probabilities of fishing operations (i.e. systematic sampling for the Norwegian Reference Fleet), a single sampling programme covering all species will not adequately address the sampling demands for all species (domains), such as spatial and temporal variations (Stock *et al.*, 2020), catchability (Rochet and Trenkel, 2005), and

Downloaded from https://academic.oup.com/icesjms/advance-article/doi/10.1093/icesjms/fsab242/6453069 by Fiskeridirektoratet. Biblioteket. user on 07 January 2022

sorting or reporting behaviour (Pitcher *et al.*, 2002). However, our results suggest that excessively large overestimations are limited to a small number of rare or non-commercial bycatch species (Figure 4). The practical consequences of non-probabilistic sampling may vary for any specific parameters observed, and across species. For example, a small spatial bias can be very significant for very localized populations if not adequately accounted for in the chosen estimator (Cosandey-Godin *et al.*, 2014).

Choice of estimator

The interpretation of our results must be viewed through the lens of the estimators we applied. We applied simple estimators as they are currently the standard practice for estimating discards and bycatches using data from the Norwegian Reference Fleet. However, reported catches tended to be overestimated even when vessels were sampled probabilistically. The magnitude of this overestimation when using probabilistic sampling is small relative to estimates using observations from Norwegian Reference Fleet vessels. However, it is still important to consider why this overestimation is occurring, and why it is larger in the trawl fishery.

The complex hierarchical structure of fisheries sampling is often ignored to simplify estimations (Nelson, 2014), either based on assumption or from evidence that sampling levels do not contribute significantly to the total variance (Tamsett *et al.*, 1999). However, accounting for the variations between sampled vessels can improve both the accuracy and precision of estimates (Moan *et al.*, 2020; Fernandes *et al.*, 2021). Our results, therefore, indicate that ignoring the hierarchical structure of Norwegian Reference Fleet sampling could result in biased estimates in total catches. This bias may be mitigated by accounting for the hierarchical structure of sampling, namely the sampling of vessels. Adequately handling the two-stage sampling design will also allow for variance estimates that may both explain part of the perceived bias and put the bias in perspective. We found small biases in our estimations in the trawl fishery, even when using probabilistic sampling of both vessels and fishing operations. The accuracy of ratio estimators is influenced by both the sample size and the strength of correlation between catches and fishing effort (Lohr, 2010). We applied the ratio estimator as this is typically employed in studies which assume a relationship between catches and fishing effort, despite evidence of little to no relationship in many cases (Diamond, 2003; Rochet and Trenkel, 2005; Cahalan *et al.*, 2015), given that other influential variables have not been accounted for, such as vessel characteristics, or on a finer spatio-temporal scale, such as within-stratum variations or variations across trips. The random forest analysis indicated that both the correlation coefficient and sample size were not important for explaining variations in biases when estimating the reported component of catches (Figure 3). However, this may be because of the limited range in values for the correlation coefficient and sample size (Supplementary Figure S3). Our random forest model was built to explore variations in biases, so whilst some variables may not have been deemed important in explaining variations in bias, it suggests that the ratio estimator was consistently biased across all species and years.

Limitations

We must also acknowledge additional biases related to total catch sampling that could not be addressed in this study. For example, official reporting of catches is enforced through inspection and carries a risk of prosecution if inaccurate. Conversely, voluntary recording of total catches may risk under-reporting sensitive portions of the catch (measurement biases; Table 1). However, based upon the data sent by Norwegian Reference Fleet vessels, the constant dialogue between scientists and fishers, payment for sampling, and supervision, we are generally confident that data collectors in the Norwegian Reference Fleet report honestly. The willingness to report sensitive information is evident in studies, which have used Norwegian Reference Fleet data to estimate “unsustainable” and “concerning” levels of bycatches of sensitive species such as seabirds (Fangel *et al.*, 2015; Bærum *et al.*, 2019) and harbour porpoise (*Phocoena phocoena*; Moan *et al.*, 2020), which are entirely absent from the official reporting framework. Nevertheless, comparisons of Norwegian Reference Fleet data with a data source of known reliability (e.g. independent observer, remote electronic monitoring; Liggins *et al.*, 1997; Faunce, 2011) could test this assumption and improve reliability from a statistical perspective (Kraan *et al.*, 2013).

This simulation study focused on the selection of vessels and hauls which was limited by the available data from mandatory reporting of catches. However, it is also important to consider biases in the biological sampling of catches which would affect size-based estimates of unreported catches. For example, clustering of samples (such as vessels in a reference fleet) can have impacts on estimates of age composition (Aanes and Pennington, 2003; ICES, 2008), whilst haul- or trip-based sampling has impacts on accuracy of estimated size distributions (Plet-Hansen *et al.*, 2020). Such biases can be identified by comparing sampling with another data source of known reliability (e.g. Starr and Vignaux, 1997).

Improving future estimations

For estimating total catches of all species in a fishery, an overly simple evaluation of bias is not appropriate. How much bias can be tolerated is dependent on a wide range of factors including intended use, accompanying precision, and sampling limitations, all of which vary between species.

Our study aimed to identify potential biases when estimating total catches. First and foremost, vessel-related biases were present in estimates using probabilistic sampling of vessels, suggesting the importance of accounting for vessel clustering of samples. Our results also suggest that the ratio estimator may be biased across all species. The variations between species are very complex and cannot be explained by encounter rates or commercial importance. There are many other species-related factors that affect bias in estimations of total catches such as patterns in spatial distribution. Therefore, alternative estimators should also be considered for individual species where assumptions of the ratio estimator are violated (Lohr, 2010), such as with rare species where the delta lognormal estimator may be more appropriate (Pennington, 1983; Ortiz *et al.*, 2000).

Although estimator bias was not directly accounted for in the random forest models, the degree to which it affects the overall bias of estimates is evident in the model interpretation (Figure 4). First, estimator bias of simple estimators resulted in an overall small yet detectable positive bias. Particularly large overestimations of reported catches for some rare species, regardless of vessel sampling method (e.g. lesser redfish in the longline fishery; Figure 4c) show that the severity of estimator bias is complex and may impact some species more than others. Fortunately, estimator bias can be evaluated in a future study in the direct context of total catches to fully understand the severity of ignoring the hierarchical sampling design.

Downloaded from https://academic.oup.com/icesjms/advance-article/doi/10.1093/icesjms/fsab242/6453069 by Fiskeridirektoratet. Biblioteket. user on 07 January 2022

In addition to determining the most appropriate design-based estimator to reduce bias, an improvement in sampling design may also reduce bias in estimates of total catches. Our results show that whilst the expert judgement selection of Norwegian Reference Fleet vessels behaves like a simple random sample for many species (Figure 4c), this assumption does not hold for all species. Increasing incentives for vessels to apply to open tenders would improve the expert judgement selection to identify the most "typical" vessels. Furthermore, in the event of multiple eligible vessels, then a larger pool of vessels would be available for the final random selection.

This study focused on the accuracy of estimators, but it is also important to consider the precision. Assumptions behind design-based estimators are different for defining sampling variance formulae (Lohr, 2010) and resampling methods for bootstrapping of variance estimates. The vessel selection biases identified in this study highlight the limitations of a small, fixed sample of vessels. For the limited number of species where bias could be deemed excessive, it is arguably better to relax assumptions in variance estimation to give a more conservative estimate in the aim of being vaguely right rather than precisely wrong.

Data availability statement

The data underlying this article were provided by the Norwegian Directorate of Fisheries by permission. Data will be shared on request to the corresponding author with permission of the Norwegian Directorate of Fisheries.

Supplementary data

Supplementary material is available at the *ICESJMS* online version of the manuscript.

Funding

This research was funded equally by the Norwegian Institute of Marine Research and the Norwegian Directorate of Fisheries.

Acknowledgements

We thank Hans Hagen Stockhausen and Geir Blom for their helpful comments on draft versions of the manuscript.

References

Aanes, S., and Pennington, M. 2003. On estimating the age composition of the commercial catch of northeast Arctic cod from a sample of clusters. ICES Journal of Marine Science, 60: 297–303.

Ainsworth, C. H., and Pitcher, T. J. 2005. Estimating illegal, unreported and unregulated catch in British Columbia's marine fisheries. Fisheries Research, 75: 40–55.

Bærum, K. M., Anker-Nilssen, T., Christensen-Dalsgaard, S., Fangel, K., Williams, T., and Vølstad, J. H. 2019. Spatial and temporal variations in seabird bycatch: incidental bycatch in the Norwegian coastal gillnet-fishery. Plos ONE, 14: e0212786.

Bell, R. J., Gervelis, B., Chamberlain, G., and Hoey, J. 2017. Discard estimates from self-reported catch data: an example from the U.S. northeast shelf. North American Journal of Fisheries Management, 37: 1130–1144.

Benoît, H. P., and Allard, J. 2009. Can the data from at-sea observer surveys be used to make general inferences about catch composition and discards? Canadian Journal of Fisheries and Aquatic Sciences, 66: 2025–2039.

Berg, H. S. F., and Nedreaas, K. 2020. Estimering av utkast i Norsk kystfiske med garn 2012-2018. Institute of Marine Research.

Borges, L., Zuur, A. F., Rogan, E., and Officer, R. 2004. Optimum sampling levels in discard sampling programs. Canadian Journal of Fisheries and Aquatic Sciences, 61: 1918–1928.

Boulesteix, A.-L., Janitza, S., and Kruppa, J. 2012. Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. WIREs Data Mining and Knowledge Discovery 2: 493–507.10.1002/widm.1072.

Bowering, R., Storr-Paulsen, M., Tingley, G., Bjørkan, M., Vølstad, J. H., Gullestad, P., and Lorentsen, E. 2011. Evaluation of the Norwegian Reference Fleet. Institute of Marine Research, Bergen. http://hdl.handle.net/11250/116704 (last accessed 10 June 2021).

Breiman, L. 2001. Random forests. Machine Learning, 45: 5–32.

Burnham, K. P., and Anderson, D. R. 2002. Model selection and multimodel inference: A practical information-theoretic approach. Springer, New York, NY. 515pp.

Cahalan, J., and Faunce, C. 2020. Development and implementation of a fully randomized sampling design for a fishery monitoring program. Fishery Bulletin, 118: 87–99.

Cahalan, J. A., Gasper, J., Mondragon, J., and Trenkel, V. 2015. Catch estimation in the federal trawl fisheries off Alaska: a simulation approach to compare the statistical properties of three trip-specific catch estimators. Canadian Journal of Fisheries and Aquatic Sciences, 72: 1024–1036.

Clegg, T., and Williams, T. 2020. Monitoring bycatches in Norwegian fisheries - Species registered by the Norwegian Reference Fleet. Rapport fra havforskningen; 2020-8. https://hdl.handle.net/11250/2685855 (last accessed 10 June 2021).

Cosandey-Godin, A., Krainski, E. T., Worm, B., and Flemming, J. M. 2015. Applying Bayesian spatiotemporal models to fisheries bycatch in the Canadian Arctic. Canadian Journal of Fisheries and Aquatic Sciences, 72: 186–197.

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., and Lawler, J. J. 2007. Random forests for classification in ecology. Ecology, 88: 2783–2792.

Diamond, S. L. 2003. Estimation of bycatch in shrimp trawl fisheries: a comparison of estimation methods using field data and simulated data. Fishery Bulletin, 101: 484–500.

Ewell, C., Hocevar, J., Mitchell, E., Snowden, S., and Jacquet, J. 2020. An evaluation of regional fisheries management organization at-sea compliance monitoring and observer programs. Marine Policy, 115: 103842.

Fangel, K., Aas, Ø., Vølstad, J. H., Bærum, K. M., Christensen-Dalsgaard, S., Nedreaas, K., Overvik, M. *et al.* 2015. Assessing incidental bycatch of seabirds in Norwegian coastal commercial fisheries: empirical and methodological lessons. Global Ecology and Conservation, 4: 127–136.

Faunce, C. H. 2011. A comparison between industry and observer catch compositions within the Gulf of Alaska rockfish fishery. ICES Journal of Marine Science, 68: 1769–1777.

Fernandes, A. C., Oroszlányová, M., Silva, C., Azevedo, M., and Coelho, R. 2021. Investigating the representativeness of onboard sampling trips and estimation of discards based on clustering. Fisheries Research, 234: 105778.

Fox, E. W., Hill, R. A., Leibowitz, S. G., Olsen, A. R., Thornbrugh, D. J., and Weber, M. H. 2017. Assessing the accuracy and stability of variable selection methods for random forest modeling in ecology. Environmental Monitoring and Assessment, 189: 316.

Friedman, B. J. H. 2001. Greedy function approximation: a graident boosting machine. The Annals of Statistics, 29: 1189–1232.

Genuer, R., Poggi, J.-M., and Tuleau-malot, C. 2009. Variable selection using random forests. Pattern Recognition Letters, 117: 256–266.

Gezelius, S. S. 2006. Monitoring fishing mortality: compliance in Norwegian offshore fisheries. Marine Policy, 30: 462–469.

Gullestad, P., Blom, G., Bakke, G., and Bogstad, B. 2015. The "Discard ban package": experiences in efforts to improve the exploitation patterns in Norwegian fisheries. Marine Policy, 54: 1–9.

Downloaded from https://academic.oup.com/icesjms/advance-article/doi/10.1093/icesjms/fsab242/6453069 by Fiskeridirektoratet. Biblioteket. user on 07 January 2022

Hothorn, T., Hornik, K., and Zeileis, A. 2006. Unbiased recursive partitioning: a conditional inference framework. Journal of Computational and Graphical Statistics, 15: 651–674.

ICES. 2008. Report of the Workshop on Methods to Evaluate and Estimate the Accuracy of Fisheries Data used for Assessment (WKACCU). 27–30 October 2008 Bergen, Norway. ICES CM 2008\ACOM:32. REF. LRC, RMC. PGCCDBS.

Jacobsen, R. B., Wilson, D. C. K., and Ramirez-Monsalve, P. 2012. Empowerment and regulation - dilemmas in participatory fisheries science. Fish and Fisheries, 13: 291–302.

Jessen, R.J. 1978. Statistical Survey Techniques. John Wiley & Sons, New York, NY.

Jones, Z. M., and Linder, F. J. 2016. edarf: exploratory data analysis using random forests. The Journal of Open Source Software, 1: 92.

Kraan, M., Uhlmann, S., Steenbergen, J., van Helmond, A. T. M., and van Hoof, L. 2013. The optimal process of self-sampling in fisheries: lessons learned in the netherlands. Journal of Fish Biology, 83: 963–973.

Liggins, G. W., Bradley, M. J., and Kennelly, S. J. 1997. Detection of bias in observer-based estimates of retained and discarded catches from a multi species trawl fishery. Fisheries Research, 32: 133–147.

Lohr, S. L. 2010. Sampling: Design and Analysis. Chapman and Hall/CRC. 596pp.

Lordan, C., Ó Cuaig, M., Graham, N., and Rihan, D. 2011. The ups and downs of working with industry to collect fishery-dependent data: the irish experience. ICES Journal of Marine Science, 68: 1670–1678.

Mangi, S. C., Dolder, P. J., Catchpole, T. L., Rodmell, D., and de Rozarieux, N. 2015. Approaches to fully documented fisheries: practical issues and stakeholder perceptions. Fish and Fisheries, 16: 426–452.

Mangi, S. C., Smith, S., and Catchpole, T. L. 2016. Assessing the capability and willingness of skippers towards fishing industry-led data collection. Ocean and Coastal Management, 134: 11–19.

Moan, A., Skern-mauritzen, M., Vølstad, J. H., and Bjørge, A. 2020. Assessing the impact of fisheries-related mortality of harbour porpoise (*Phocoena phocoena*) caused by incidental bycatch in the dynamic Norwegian gillnet fisheries. ICES Journal of Marine Science, 77: 3039–3049.

Nelson, A. G. 2014. Cluster sampling: a pervasive, yet little recognized survey design in fisheries research. Transactions of the American Fisheries Society, 143: 926–938.

Norwegian Directorate of Fisheries. 2021. Norske Omregningsfaktorer - For omregning av landet produktvekt av marin fisk og andre marine arter til rund vekt - for fiske i det nordlige Atlanterhavet og i andre farvann. https://www.fiskeridir.no/Yrkesfiske/Tema/Omregningsfaktorer (last accessed 10 June 2021).

Ortiz, M., Laegault, C. M., and Ehrhardt, N. M. 2000. An alternative method for estimating bycatch from the U.S. shrimp trawl fi shery in the Gulf of Mexico, 1972–1995. Fishery Bulletin, 98: 583–599.

Pennington, M. 1983. Efficient estimators of abundance for fish and plankton surveys. Biometrics, 39: 281–286.

Pitcher, T. J., Watson, R., Forrest, R., Þór Valtýsson, H., and Guénette, S. 2002. Estimating illegal and unreported catches from marine ecosystems: a basis for change. Fish and Fisheries, 3: 317–339.

Plet-Hansen, K. S., Bastardie, F., and Ulrich, C. 2020. The value of commercial fish size distribution recorded at haul by haul compared to trip by trip. ICES Journal of Marine Science, 77: 2729–2740.

Rochet, M.-J., and Trenkel, V. M. 2005. Factors for the variability of discards: assumptions and field evidence. Canadian Journal of Fisheries and Aquatic Sciences, 62: 224–235.

Roman, S., Jacobson, N., and Cadrin, S. X. 2011. Assessing the reliability of fisher self-sampling programs. North American Journal of Fisheries Management, 31: 165–175.

Siders, Z., Ducharme-Barth, N., Carvalho, F., Kobayashi, D., Martin, S., Raynor, J., Jones, T. *et al.* 2020. Ensemble random forests as a tool for modeling rare occurrences. Endangered Species Research, 43: 183–197.

Snyder, H. T., and Erbaugh, J. T. 2020. Fishery observers address arctic fishery discards. Environmental Research Letters, 15: 0940c4.

Starr, P. 2010. Fisher-Collected sampling data: lessons from the New Zealand experience. Marine and Coastal Fisheries, 2: 47–59.

Starr, P. J., and Vignaux, M. 1997. Comparison of data from voluntary logbook and research catch-sampling programmes in the New Zealand lobster fishery. Marine and Freshwater Research, 48: 1075–1080.

Stock, B. C., Ward, E. J., Eguchi, T., Jannot, J. E., Thorson, J. T., Feist, B. E., and Semmens, B. X. 2020. Comparing predictions of fisheries bycatch using multiple spatiotemporal species distribution model frameworks. Canadian Journal of Fisheries and Aquatic Sciences, 77: 146–163.

Stratoudakis, Y., Fryer, R. J., Cook, R. M., and Pierce, G. J. 1999. Fish discarded from Scottish demersal vessels: estimators of total discards and annual estimates for targeted gadoids. ICES Journal of Marine Science, 56: 592–605.

Strobl, C., Boulesteix, A. L., Zeileis, A., and Hothorn, T. 2007. Bias in random forest variable importance measures: illustrations, sources and a solution. BMC Bioinformatics, 8: 25.

Strobl, C., Boulesteix, A., Kneib, T., Augustin, T., and Zeileis, A. 2008. Conditional variable importance for random forests. BMC Bioinformatics, 11: 307.

Strobl, C., Hothorn, T., and Zeileis, A. 2009. Party on! A new, conditional variable importance measure available in the party package. The R Journal, 1: 14–17.

Suuronen, P., and Gilman, E. 2020. Monitoring and managing fisheries discards: new technologies and approaches. Marine Policy, 116: 103554.

Tamsett, D., Janacek, G., Emberton, M., Lart, B., and Course, G. 1999. Onboard sampling for measuring discards in commercial fishing based on multilevel modelling of measurements in the irish sea from NW England and N Wales. Fisheries Research, 42: 117–126.

Vølstad, J. H., and Fogarty, M. 2006. Report on the National Observer Program vessel selection bias workshop Woods Hole, May 17-19, 2006.

Handling Editor: Ernesto Jardim

Downloaded from https://academic.oup.com/icesjms/advance-article/doi/10.1093/icesjms/fsab242/6453069 by Fiskeridirektoratet. Biblioteket. user on 07 January 2022