FISKERIDIREKTORATET

Rapport

Estimating the size distribution of reported catches on-board factory vessels – Issues with using data from the production process

Thomas L Clegg, Geir Blom, Kotaro Ono, Kjell Nedreaas, oktober 2021

Livet i havet – vårt felles ansvar

Rapport

Estimating the size distribution of reported catches on-board factory vessels – Issues with using data from the production process

Årstall

2021

Ansvarlig avdeling:

Ressursavdelingen og Statistikkavdelingen (Fiskeridirektoratet)/ Fiskeridynamikk (Havforskningsinstituttet)

Emneord:

High-grading, bycatch, discard, fishmeal, factory vessels, length-distribution, bias, precision

Arkivsaksnummer:

Dato utgitt:

01.11.2021

Totalt antall sider:

22

Saksansvarlig:

Thomas L. Clegg, Geir Blom, Kotaro Ono og Kjell Nedreaas

ISBN: ISBN-13: 9788292075098

Sammendrag

‘High-grading’ er praksisen med å kaste ut fangster med lavere verdi for å få plass til fangster med høyere verdi. Det er nødvendig å forstå omfanget av urapportert utkast og dets variasjon for å forbedre bestandsvurderinger og forvaltningsbeslutninger. Der det ikke er direkte observasjoner av utkast av fisk, finnes det metoder for å estimere ‘high-grading’ ved å sammenligne størrelsesfordelinger av totalfangst (før sortering) og landet fangst (etter sortering), men per i dag er det ikke tilgjengelig en tilstrekkelig detaljert datakilde for den landete delen. Denne rapporten presenterer data fra to pilotstudier for å undersøke egnetheten til data samlet inn i forbindelse med om bord produksjon for å beskrive størrelsesfordelingen av rapporterte fangster. Produksjonsrapportene vi mottok inneholdt aggregerte data, der individuelle vekter av en fiskeart var aggregert i grove, overlappende vektintervall. Slike data resulterer i et stort tap av informasjon. Bruk av statistiske prosedyrer for å få et mer detaljert bilde av størrelsesfordelingene av fisk, vil introdusere enda mer usikkerhet i et allerede usikkert datamateriale som potensielt kan føre til ikke-signfikante resultater og som kan introdusere ukjente skjevheter. Vi konkluderer med at det er nødvendig å bruke rådataene bak aggregerte produksjonsrapporter. Rådataene består av registrert produktvekt av ulike fiskearter på individnivå. I denne forbindelsen diskuterer vi de logistiske og statistiske problemene som produksjonsdata inkludert observasjoner på individnivå kan introdusere.

Summary

High-grading is the practice of discarding lower value catches to make space for catches with higher value. It is necessary to understand the extent and variation in these unreported discards to improve stock assessments and management decisions. Where discards are not directly observed, a proposed methodology for estimating high-grading involves comparing size distributions of total catches (before sorting) and landed catches (after sorting), but we have yet to identify a suitable data source for the landed portion. This report presents data from two pilot studies exploring the suitability of data gathered during the on-board factory production process for describing the size distribution of reported catches. We received these data in a summarised report, where individual fish weights are aggregated into coarse, overlapping size grades. This summarised form results in a large loss of information. Applying the necessary statistical procedures to get a more detailed picture of fish size distributions would introduce even more uncertainty into an already uncertain estimation, potentially leading to non-signficant results, and can introduce unknown biases. We conclude that it is necessary to use the raw data behind summarised reports which provide data on individual fish. To this end, we address the logistical and statistical issues posed by production data including individual fish observations.

Rapport

Estimating the size distribution of reported catches on-board factory vessels

Innholdsfortegnelse

Tables

Figures

1. Introduction

High-grading is the practice of discarding lower value catches to make space for catches with higher value (Batsleer *et al.*, 2015). High-grading of commercial species is typically size-based, influenced by the minimum landing size or market prices favouring larger individuals. Discarding is illegal under a landing obligation and therefore results in misleading catch statistics. For example, neglecting discards of small fish in stock assessments can mask strong incoming year-classes (Punt *et al.*, 2006). A good knowledge of high-grading is therefore necessary to improve stock assessments (Dickey-Collas *et al.*, 2007; Perretti *et al.*, 2020) and have a more realistic understanding of the environmental impact of the fishery.

The Norwegian discard ban was implemented in 1987 to mitigate against the emerging practice of high-grading (Gullestad *et al.*, 2015). Since then, a suite of accompanying measures has been developed, known collectively as the 'discard ban package' (Gullestad *et al.*, 2015), to build a more comprehensive policy to reduce discards. Fishers are incentivised to land illegal catches through compensation; avoidance is actively encouraged through legal obligations to move away from high-risk areas; and fishing gears are being constantly developed to improve selectivity. However, it is known that discarding still occurs through direct observations by the Norwegian Coast Guard that have resulted in prosecutions. There are also various studies estimating historical high-grading in the Barents Sea trawl fishery for cod (*Gadus morhua*). In the absence of direct observations, empirical gear selectivity curves have been applied to fisheries-independent sampling to simulate commercial fishing (McBride and Fotland, 1996; Dingsør, 2001). Breivik *et al.* (2017) estimated historical bycatches of cod in the Barents Sea shrimp fishery using the Directorate of Fisheries Monitoring and Surveillance Service (MSS), a programme which hires or joins fishing vessels. However, this sampling is focussed on species of concern (e.g. juvenile cod bycatch) and does not regularly record size measurements of species.

More recently, the Norwegian Reference Fleet have provided direct sampling of species (Clegg and Williams, 2020), including size measurements which allows for more direct methods for quantifying high-grading. In the coastal segment of the Norwegian Reference Fleet, vessels sample discarded and landed portions of the catch separately, which enables high-grading to be readily identified (Berg and Nedreaas, 2020). However, in offshore fisheries, the Norwegian Reference Fleet have only sampled discards and landed catches separately since 2019. Prior to this, vessels only sampled total catches (i.e. discards and landed catches combined). In this situation, we can still estimate fishery-level high-grading by comparing the size distributions of observed total catches with those of landed catches (e.g. Pálsson, 2003). However, we have yet to identify a reliable source of size-based data on landed catches to make such a comparison.

There are numerous size-based data sources in the mandatory catch reporting framework, but unlike the Norwegian Reference Fleet, official catch reports do not offer an adequate data resolution for quantifying high-grading. Daily logbooks have a high spatial and temporal resolution, but do not record size-based information on catches. Sales notes are generated once a vessel lands the catches after each trip. They do include size-based information, but only as a summary of an entire trip and in course market-defined size grades. In one trip, a vessel spans many statistical areas over a period of weeks or even months, meaning trip-level resolution is insufficient to understand spatial and temporal patterns in high-grading.

For offshore fisheries prior to 2019, where discards were not reported explicitly, the other scientific sampling programmes did not provide information on unreported portions of the catch. Coast Guard sampling is done by enforcement officers, meaning that fishers will not discard in their presence. A port intercept sampling programme only samples coastal vessels landing fresh fish north of 62°N latitude (Hirst *et al.*, 2004), whilst a newer mandatory self-sampling programme covers offshore fisheries for selected pelagic species (Stenevik *et al.*, 2020).

This report focuses on estimating high-grading in offshore Norwegian fisheries prior to 2019, where the Norwegian Reference Fleet did not sample landed and discarded catches separately. We present two pilot studies which trialled a potential source of size-based data of reported catches generated by on-board factory production systems. Vessels constantly monitor productivity in the on-board factory to inform fishing strategy and to keep a record of catches on board, which also contributes to mandatory reporting. At present, these data are the most detailed source of size-based information for retained catches. We requested summarised reports (hereafter referred to as production reports) from vessels in two fisheries to explore their viability for use in quantifying high-grading. We describe the production process on-board factory vessels and how production reports are generated. We then explore the statistical properties of the data to identify issues that restrict their viability for quantifying high-grading and discuss how these issues can be rectified for future studies.

2. Description of the production process on board factory vessels

As fish enter the on-board factory immediately after hauling, crew sort through the catches to decide which fish to process (Figure 1). Fish can either be processed into a range of products, discarded, or converted into highly-processed products such as fishmeal or ensilage.

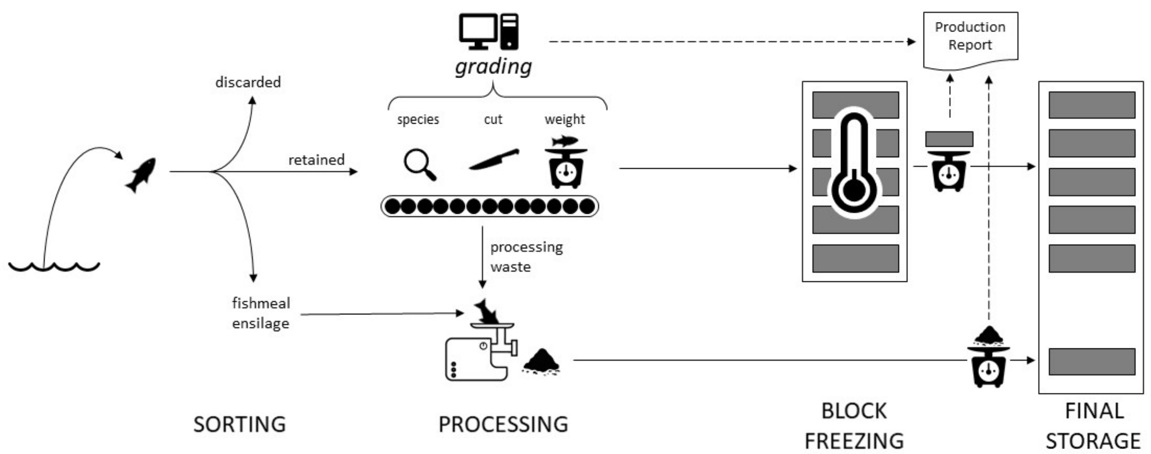

Figure 1: Generalised description of the production process on-board a factory vessel. The specific process can vary depending on factory setup.

During the production process, individual fish are cut for the desired product and then weighed. This means there is no information on the round weight (original live weight) of fishes. The species is either registered manually, inferred from the route taken through the factory, or more recently by image recognition software. Using automatic conveyor belt systems, all this information is combined to grade each fish based on the species, product and processed weight. Graded products are held in temporary storage tanks, where they are then frozen in standardised blocks and transferred to the final storage hold.

The production process generates no information about unreported catches, which includes illegally discarded fish and fish processed as fishmeal (NOU, 2019). We therefore cannot directly quantify the scale of high-grading due to illegal discarding, and how much is legally landed as fishmeal. However, if a vessel has on-board fishmeal production, then we can assume that all unwanted fish are processed into fishmeal, meaning unreported catches on those vessels are not a result of high-grading. The majority of vessels do not have fishmeal production facilities on-board, but new vessels are increasingly installing them. If the vessel cannot produce fishmeal, unwanted catches can be frozen whole as mixed species to be delivered to production facilities on land.

To monitor production and assist in mandatory reporting, summary reports can be generated for any given time period. The report aggregates the number and weight (measured or estimated) of frozen blocks from the final output stage to provide the total weight of each product and grade. The report can also be supplemented with additional information from the grading process, such as the average individual weight in each grade, which is used for estimating the total number of pieces of each product.

If the factory exceeds production capacity, crew can bypass the automated steps in the process to speed up production. This can result in certain information not being recorded. For example, we observed one system where crew record species and product by dropping incoming fish into defined hoppers. The hoppers then drop fish onto a conveyor belt at defined intervals, which passes over a weighing scale to record its weight. If a catch is large and dominated by a certain species and grade, the defined hopper may not drop fish fast enough, causing a backlog in the system. Crew can avoid this by storing those fish in baskets, then manually adding them to the correct temporary storage tanks after production has calmed, knowing the weight will be registered after freezing. This solution creates a risk that not all fish are recorded in the grading machine. Furthermore, these manually graded fish will likely differ in size to those automatically graded.

Using production reports as a source of information on high-grading assumes that all fish entering the processing stage are ultimately landed and reported. Vessels are legally required to maintain an accurate record of all catches stored on the vessel. These catch diaries are filled out based upon production data described above. The catch diaries can be inspected at any time at sea by enforcement authorities. They should then match with the sales of catches after returning to land. Substantial inconsistencies between the catch diary and landing report will be automatically flagged and can be investigated. Norway has previously been ranked highest in fisheries compliance globally (Pitcher *et al.*, 2009), owing to efficient regulations and enforcement, and a broad willingness for fishers to comply (Gezelius, 2006). Therefore, for the purposes of a study on high-grading, we can assume that production reports are representative of reported catches with respect to reliability of reporting.

3. Pilot studies – production report data requests

We ran two pilot studies to investigate the utility of production report data for estimating size distributions of species in reported catches in offshore fisheries. For both pilot studies, we contacted a selection of vessels to voluntarily provide daily production reports. The letter stated that data will be used for estimating unreported bycatches in the fishery, and that raw data will be treated as sensitive in accordance to data privacy laws. The letter also stated that vessels will not be prosecuted on the basis of submitted data, and that published materials will be aggregated and anonymised such that individual vessels cannot be identified.

The first pilot study was the Barents Sea trawl and autoline fisheries (Figure 2) for cod (*Gadus morhua*) and haddock (*Melanogrammus aeglefinus*) in 2012. We randomly selected 10 trawl vessels and five autoline vessels for each annual quarter, with a probability weighted by total reported catches in the previous year. The selection for each annual quarter was independent, such that it was possible to select a vessel for multiple quarters. We contacted a total of 45 vessels, of which 18 vessels cooperated by providing the requested data. In four cases, reports were summarised over periods longer than a day, ranging from 14-25 days.

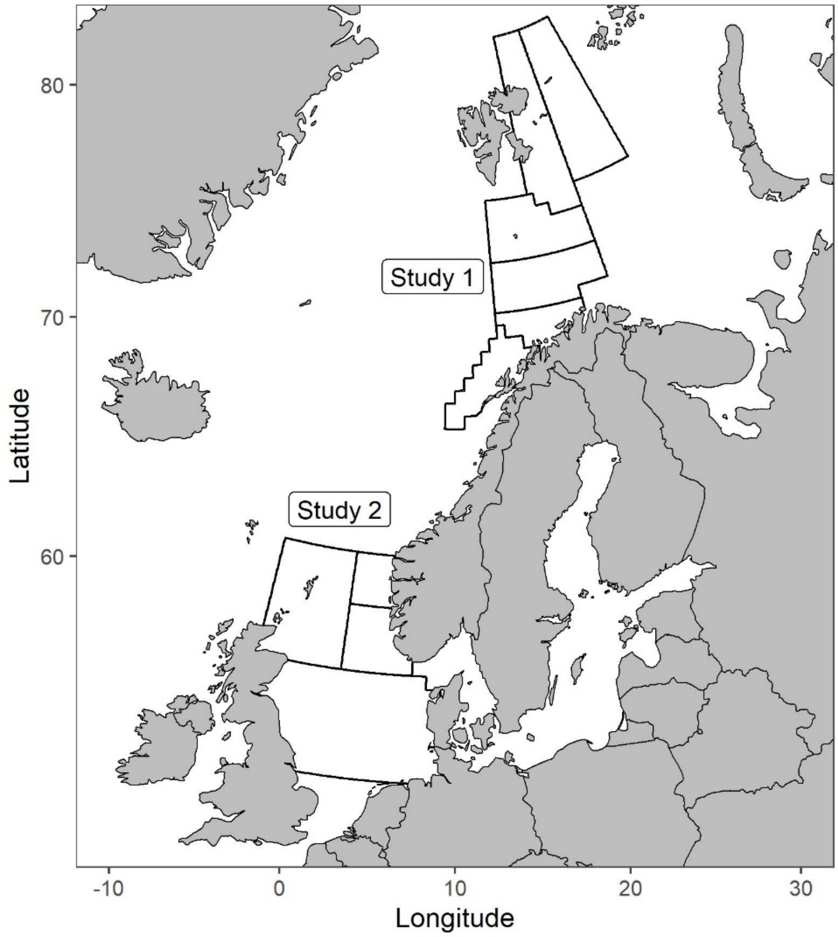

Figure 2: Map of study fisheries showing statistical areas included in the two pilot studies.

The second pilot study was the North Sea trawl fishery for saithe (*Pollachius virens*) (Figure 2) in 2018. We replicated the request from the first pilot study, with 13 vessels selected for each annual quarter, again with the possibility of multiple requests across quarters. We additionally requested all vessels to provide supplementary information from the grading machine for each daily production report, namely the mean weight of individual fish in each size grade. Of the 31 vessels contacted, 14 responded with production report data. All vessels returned reports for individual days. Only three vessels provided the supplementary information on mean weight from the grading machine data, which were summarised over periods ranging from one to 32 days.

We additionally requested four vessels to provide the raw data from the grading machine that is used to generate the supplementary information on mean weight. However, no vessels fulfilled this request.

3.1. Issues with data requests

In both pilot studies, data were typically provided as printable reports structured to improve readability, which hindered data entry. Furthermore, many vessels provided reports in PDF or paper formats, requiring an additional step to digitise and extract the values.

A large number of companies did not respond to our request. Unless a company actively objected to the request, we cannot determine if the rejection was intentional objection or due to neglect. This is important to understand due to the implications on sampling biases. If companies object to the request for similar reasons (e.g. high risk of prosecution), and furthermore, those characteristics differ from the general population, then the final sample may not be representative of the fishery. The request required a large administrative task to access, filter, and compile the dataset, then send it. For this reason, it is possible that many of the non-responses may be neglectful and could therefore be deemed random and not introduce bias.

4. Estimating unreported catches using production reports

To identify high-grading in a fishery where we do not have direct observations of discarded catches, we can infer it by comparing the size distribution of total catches with that of landed catches. If the two size distributions are different, then we attribute this to high-grading (Pálsson, 2003). It is important to highlight that we cannot determine if these unreported catches were discarded or converted to fishmeal.

Size distributions of total catches are available from the Norwegian Reference Fleet catch sampling programme. Vessels participating in the programme regularly record the length and weight of fish from samples of total catches (see Clegg and Williams (2020) for detailed sampling protocols), which can be aggregated across vessels to give a high-resolution size distribution of species for any given temporal or spatial scale.

In their standard format, production report data are not comparable to the size distributions generated from individual fish measurements by the Norwegian Reference Fleet. Production reports summarise total weight of products for each species in different size grades (Table 1). Therefore, a comparison would require the transformation of one or both datasets to standardise them.

One solution would be to reduce the resolution of Norwegian Reference Fleet data by aggregating individual fishes into size grades that match production reports. Even ignoring the complications in this process, reducing data resolution is highly undesirable in principle. It would be wasteful for the effort and money spent in sampling individual fishes and would only increase uncertainty.

Alternatively, we could infer the underlying size distribution of species from aggregated production reports. This would offer an estimated number of fish in smaller size intervals comparable with Norwegian Reference Fleet sampling. In this section, we present potential methods for inferring underlying size distributions, and explain why the data structure of production reports creates issues with this approach.

4.1. Round weight vs processed weight

The first obstacle met is the difference in measures of weight between processed catches in production reports and total catches recorded by the Norwegian Reference Fleet. Where fish are processed on-board the vessel, then fish weights are recorded after processing and

Table 1: Extract from a daily production report listing total production in a 24-hour period in January 2012 in statistical area 12.

Species	Product	Grade (kg)	Total product weight (kg)
Cod *Gadus morhua*	Fillet with bone with skin	0.45–0.91	1222
		mix	1364
	Headed & gutted	1–2.5	1749
		2.5–5	2325
		>5	144
Golden redfish *Sebastes norvegicus*	Headed & gutted	>1	79
	Headed & gutted (Japan cut)	≤1	47
Haddock *Melanogrammus aeglefinus*	Fillet without bone with skin	mix	819
	Headed & gutted	≤0.8	1260
		>0.8	1127
Halibut *Hippoglossus hippoglossus*	Headed & gutted	≤6	15
Spotted wolffish *Anarchichas minor*	Headed & gutted	≤3	43
Tusk *Brosme brosme*	Headed & gutted	≤1	36

converted back into round weight (the weight of the fish when it is taken from the water) using official conversion factors calculated by the Norwegian Directorate of Fisheries (Norwegian Directorate of Fisheries, 2021). The conversion factors are based on sampling on-board active fishing vessels, where fish are weighed before and after processing to estimate the average weight lost from production for each product. Conversion factors are published as annual mean values for all areas and are published without estimation error and are updated intermittently.

4.2. Weight-based vs. length-based assessment

Size-based stock assessments typically structure fish populations by age or length. The Norwegian Reference Fleet gather data on weight, length and age of individual fishes, providing usable data for stock assessments. Comparatively, production reports only include fish weight and would therefore require a conversion to either length or age to be comparable to data on total catches and subsequently useful for stock assessments.

4.3. Parametric approach

A parametric approach to inferring an underlying size distribution involves an assumption that observations come from a known distribution, which is described by a fixed set of parameters. Simpler distribution fitting methods in R (e.g. *MASS::fitdistr*; Venables and Ripley, 2002) assume that observations are known without error. This is not true for observations from production reports, which only offer the total weight of fish in a defined size range (Table 1). Fortunately, more advanced methods of distribution fitting can account for such uncertainties, such as the *fitdistrplus* package in R (Delignette-Muller and Dutang, 2015). This expands the distribution fitting functions to accommodate both interval and censored observations, which are key characteristics of production reports (Table 1).

4.4. Interval and censored size grades

In production reports, fish size grades are reported in intervals, such that we only know that fish were within a defined weight range. The largest and smallest size grades are typically censored, meaning that the size range is only partially known. Left-censored size grades (i.e. fish below a certain size) are limited to positive values, which is reflected in the appropriate distributions for weight data, namely gamma and log-normal distributions. However, these distributions can include values approaching zero. It is extremely unlikely to observe fish with a size approaching zero in catches due to size selective fishing gears and avoidance strategies (Reid *et al.*, 2019).

Right-censored size grades (i.e. fish above a certain size) include the largest fish caught in the fishery. The most likely reasons for unreported large fishes are either discarding of damaged individuals, or illegal sale. A right-hand censored size grade has no theoretical limit so if left undefined, the distribution fitting functions will estimate the limit based on the estimated parameter for the distribution. This limit could also be defined empirically, based on the largest fish observed in the Norwegian Reference Fleet if it improved model fit quality or convergence.

The *fitdistrplus* package in R (Delignette-Muller and Dutang, 2015) can account for censored size grades by estimating the cumulative distribution function for censored observations instead of probability density function for non-censored observations. Assessing the quality of model fit is not a simple comparison, but instead requires a judgement based on a suite of statistical tests and the visual inspection of graphical outputs. This is further complicated by the reason for fitting a distribution. In our situation, we want to identify if high-grading is occurring, which we expect to be size-based, such that small fish are more likely to be misreported. Any differences in size distributions between total and reported catches are therefore more likely in the left tail of the distribution. Fortunately, assessing the quality of distribution fitting can focus on this portion of the distribution at the expense of the right tail (larger fish) (Delignette-Muller and Dutang, 2015).

In extreme cases, a portion of catches may not be graded at all. This can be a result of market demands, grading errors, or damaged products. We have no information on whether this grade allocation is biased with respect to size, but an assumption of no bias will allow mixed grades to be removed from an analysis.

4.5. Estimating the number of fish in each size grade

To fit a distribution to fish size observations using censored data, we must first know the number of fish in each size grade. If only the total catch weight is reported for each size grade, we must estimate the number of fish by dividing the total weight by the average individual weight in each grade. An estimation of numbers of fish in each size grade based upon empirical data is desirable over using the midpoint of each size grade.

In the absence of an empirical estimate of the average individual weight, we must assume it to be the midpoint of each size grade. This assumes that size grades are independent of each other, and that in each size grade, observed weights are normally distributed and centred around the midpoint. In reality, size grades are not independent; they are an arbitrary division of a larger size distribution. Size grades in the left tail of the distribution will contain increasing numbers of larger fish, whilst those in the right tail will contain decreasing numbers, meaning that fish sizes in one size grade are seldom normally distributed. Considering in addition that larger fish contribute more weight, it is likely that the midpoint is a poor estimator of average individual fish weight in each size grade.

In the first pilot study, we did not have any empirical knowledge of average individual weight, so we must assume that it is the midpoint of each size grade. In the second study, some vessels provided the mean individual weight of fish in supplementary data from the grading machine. Using these additional data from the second study, we can demonstrate the importance of using empirical estimates of average individual weight. We first calculate the number of fish in each size grade using both the assumed average individual size (midpoint) and the empirical estimate (supplementary grading data). We can then calculate the percentage error introduced from assuming the average individual size:

$$Error_{j,k}\ (\%) = \frac{1}{N}\sum_{i=1}^{N}\frac{\hat{x}_{ijk} - x_{ijk}}{x_{ijk}} \times 100 \tag{1}$$

where for an individual production report, i for species, j and size grade, k, $\hat{x}_{ijk}$ is the assumed number of fish based upon the midpoint of the size grade and x_{ijk} is the estimated number of fish taken from supplementary data. For right-censored size grades, we assumed the upper limit was the largest fish observed by the Norwegian Reference Fleet.

Figure 3 shows that assuming the average individual weight of fish in each size grade is the midpoint of that grade results in large biases in estimates of total number of fish for all size grades. Furthermore, there is a strong trend across all three species to overestimate the number of fish in left-censored size grades when using the midpoint method. An overestimation of small fish in reported catches will mask the true scale of high-grading when comparing with observations of total catches by the Norwegian Reference Fleet. This issue is worsened by the underestimation of numbers of larger fishes. When fitting a parametric distribution to the size data, an over- and underestimation of small and large fish respectively will force the best-fitting distribution to be more positively skewed. Given that we expect discarding to be relatively low in Norwegian fisheries (Pérez Roda *et al.*, 2019; Gilman *et al.*, 2020), there is an increased risk that an underestimation of high-grading could conclude that it does not occur.

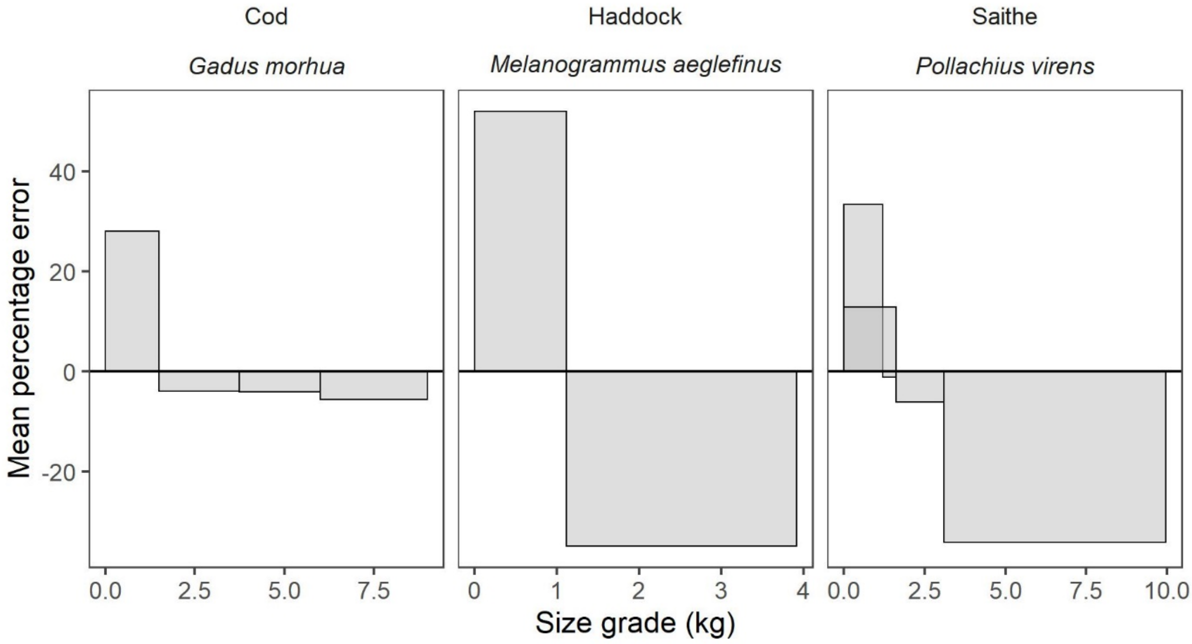

Figure 3. Error in estimated number of fish from assuming the average individual weight is the midpoint of each size grade. Each bar represents a size grade, which may overlap (e.g. Saithe). Left- and right-most grades are censored. Censored size grades were given assumed limits to allow for plotting (left-censored = zero; right-censored = largest individual observed by the Norwegian Reference Fleet.

4.6. Variations in grading intervals

There is no standardisation of grading intervals across the fishing fleet, as intervals vary depending on business strategy, market trends and catch composition. This makes it impossible to aggregate production reports on any level for many species without avoiding overlapping size grades. We could select only those reports with matching grading intervals to create a reduced dataset which could be aggregated. However, we do not know enough about the detailed reasons for grading decisions to understand whether removing certain grading intervals would bias sampling.

4.7. Coarse grading

For some species such as haddock, grading is limited to a small number of categories. For example, across both pilot studies 58.1% of daily production reports for haddock used only two size grades. Furthermore, both of these grades are censored (i.e. either larger or smaller than a defined weight). Similarly, 16.4% of daily production reports contained only one grade for either cod or haddock. In these situations, there is simply not enough information in the summarised reports to estimate the underlying distribution.

With wide, left-censored size grades, it is possible that only one size grade describes the left tail of the distribution. In this situation, it is likely that the distribution of fish sizes will be best described as an exponential decrease. Candidate distributions that typically describe fish sizes, such as the gamma and log-normal distributions, can approach or become exponential given certain parameterisations. An exponential distribution is not suitable for describing the size distribution of fish caught by trawl and autoline fishing gears, which are selective for larger individuals (Reid *et al.*, 2019).

4.8. Multimodal size distributions

We cannot determine if a multimodal distribution should be fitted based on the information from production reports (Table 1), due to grading being very coarse and censored. Fitting unimodal (single-peaked) distributions to infer the underlying fish size distribution ignores the possibility of multimodal size distributions (multiple peaks). Multimodal size distributions are common in fish populations due to a wide range of biotic and abiotic factors (Huston and Deangelis, 1987). These occur either as multiple factors influencing the same population, or the unintentional combining of distinct groups of fishes (e.g. populations or different life history stages). Fish are typically seasonal spawners, meaning they are born in a discrete time period and grow as a cohort. However, individual growth rates will vary due to a wide range of factors such as genetics, environment and food availability (Huston and Deangelis, 1987). Habitat associations at different life stages can separate fish size on a fine spatial scale and cause seasonal variations in size distributions (Methratta and Link, 2007). It is difficult to capture these factors for all species using a single stratification system.

Figure 4 shows two examples of multimodal size distributions in total catches sampled from both Norwegian Reference Fleet and Coast Guard inspections. In such examples, we could be observing multiple cohorts, sex-dependent growth differences or overlapping populations. Whilst we can speculate any number of underlying causes here, the main point is that fitting a unimodal distribution to these two examples would skew the peak towards the major modes. Furthermore, this issue is focused around the minimum landing size of the species, which is the focal area for quantifying high-grading.

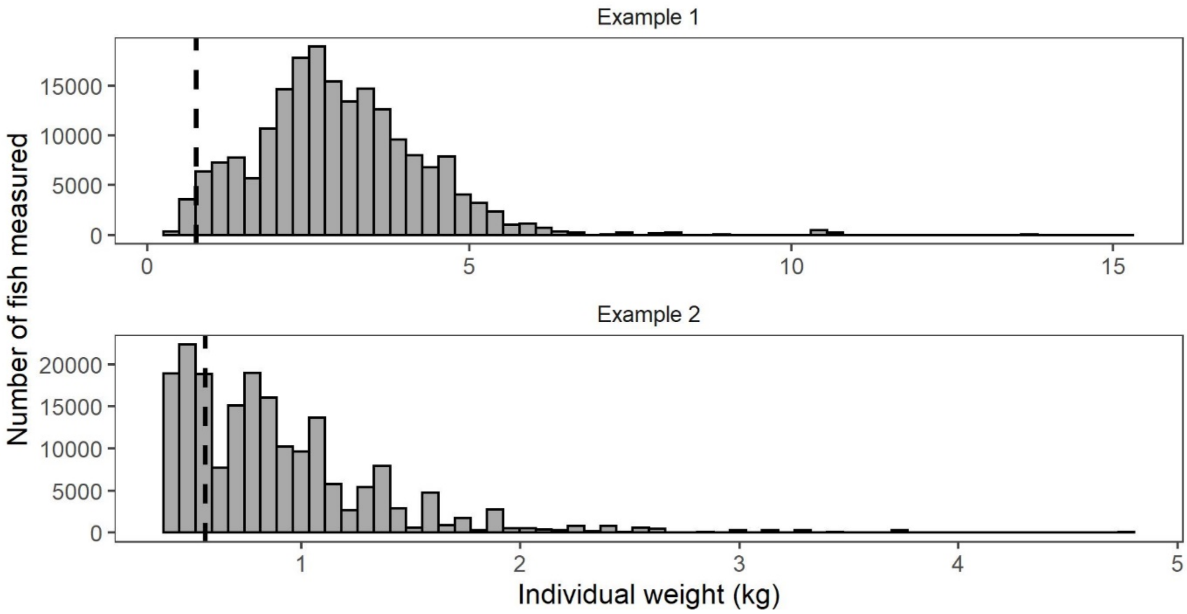

Figure 4. Examples of multimodal fish size distributions using data from the Norwegian Reference Fleet sampling and Coast Guard inspections. Dashed line indicates minimum landing size (converted from length). Example 1: Cod in the Barents Sea trawl fishery, statistical area 23, quarter 4; Example 2: Saithe in the North Sea trawl fishery, statistical area 08, quarter 3.

4.9. Non-parametric approach

Parametric approaches lack the ability to describe more complex distributions for which we do not know the underlying assumptions, the largest of these being multimodal distributions. On the other hand, non-parametric models make no assumptions about the distribution of data. As a result, non-parametric models need more data to understand the underlying functions, or more importantly, they need more information.

A non-parametric approach to estimating the underlying size distribution involves the ungrouping of coarsely aggregated data into smaller intervals (Rizzi *et al.*, 2015). A comparison of methods for ungrouping coarsely aggregated data by Rizzi *et al.* (2016) identified the penalized composite link model as the most efficient method for very coarsely aggregated data. The penalized composite link model can account for censored size grades but requires an estimation of the size of the right-censored grades (left-censored grades are limited at zero). In epidemiological applications, a right-censored age group is more easily limited to the oldest known age (Rizzi *et al.*, 2016), which is relatively easy to estimate. However, it is more difficult to estimate the largest fish in the fished population, as we only have information on those fish caught in samples rather than census information.

The largest obstacle to using a penalized composite link model for estimating the underlying size distribution is the need for sequential intervals in the raw count data. However, fish are graded depending on a wide range of factors including product, market demand and catch composition. Therefore, even on the level of a single haul or product, there is seldom a sequential grading system (Table 1). Even if a sequential grading system was used for each haul, product or time period, we would need to fit individual models to each haul, product or time period independently, reducing the data available for analysis and increasing risk of uncertainty and bias in estimates when combining the model outputs.

5. Discussion

This report has explored the statistical properties of data generated by production reports on board factory fishing vessels, which were obtained through two pilot studies to determine their utility for quantifying high-grading in offshore fisheries where scientific sampling of reported catches is unavailable.

5.1. Data collection

A large number of companies failed to provide production reports upon request, which could result in non-response biases. There were some cases of apprehension towards the intended use of the data, despite a clear definition of intentions and reassurance of safe data handling protocols. Some companies offered that independent on-board observers could collect the required data, but such sampling programmes are time-consuming and expensive. Lohr (2010) stated that by far the best method of dealing with non-response is its prevention. In the context of this study, prevention would involve building trust with fishers or incentivising cooperation such that the risk of non-response would be low and less biased. There is a legal obligation to provide such data upon request, which if enforced would result in 100% response rate. However, our experiences in the two pilot studies highlighted some possible negative impacts of mandatory provision. Unreported catches are a sensitive issue and scientific results will have direct consequences to control and management of the fisheries. A willingness to contribute data can increase the acceptance of results (Hoare *et al.*, 2011; Mangi *et al.*, 2016), so fostering a trustful cooperation with the fishing industry at the earliest possible stage will improve the long-term success of monitoring of unreported catches.

If non-response cannot be prevented, then an understanding of the statistical properties of the non-respondents would reveal potential biases. Fortunately, daily logbooks, sales notes data and vessel GPS tracking provide highly detailed information on the characteristics of offshore vessels. Taking a representative sample of non-respondents would allow for inference about other non-respondents (Lohr, 2010).

Considerable efforts were taken by companies to provide production reports and by research staff to standardise and compile the data. Nevertheless, some reports covered the wrong time period and areas, or were summarised at the wrong resolution. The excessive time spent in processing data and rectifying issues suggests that the data collection methodology is not scalable to other case studies or over multiple years.

5.2. Comparison with total catch sampling to identify high-grading patterns

5.2.1. Information loss

We base our analysis on the assumption that any differences between size distributions of total and landed catches can be attributed to high-grading. However, that assumption breaks down as we must further process data before making a comparison. Firstly, the statistical methods may introduce biases which cannot be quantified, but which may affect the interpretation of results. Secondly, uncertainties introduced by the data processing methods will likely increase uncertainty, making it more difficult to identify small effect sizes that come with low levels of high-grading. This is an important consideration given that it is commonly assumed that Norway has low levels of discarding (Pérez Roda *et al.*, 2019; Gilman *et al.*, 2020) due to high compliance and a well-established discard ban.

Supplementing production report information with summarised data from the grading machine (average product weight) may alleviate the issue of estimating the number of fish caught (NOU, 2019), but it does not address the issues of coarse grading systems, censored size grades, and the possibility of multimodal distributions. The only way to alleviate these issues using production report data would be to use the individual fish measurements from the graders.

5.2.2. Converting product weight to length

The lack of observations of fish lengths in on-board factories requires a two-stage conversion before analysis. Official conversion factors for round weight by the Norwegian Directorate of Fisheries are presented for each product as single annual values with no measure of uncertainty. In reality, product yield varies depending on the quality of the cut and size of the fish, the latter of which varies both spatially and seasonally (Mello and Rose, 2005). Converting round weight to length or age is further necessary for results to be suitable for input into stock assessments. Both these conversions will introduce uncertainty (and possibly bias), which should be accounted for in final estimates of high-grading. Ignoring this uncertainty could risk a type I statistical error where significantly important levels of high-grading are reported due to misleadingly small uncertainties in the estimate.

Uncertainty in product weight to round weight conversions are available for selected species and products (Blom, 2014), whilst weight-length data are available from Norwegian Reference Fleet sampling to estimate the relationship. Both these sources of uncertainty can be factored into estimations using Monte Carlo simulations. This involves repeatedly generating random observations from within the known range of uncertainty to produce a full range of possible outcomes for the final estimate.

In the absence of sufficient information on uncertainty, a sensitivity study would help to understand the level of uncertainty necessary to cast doubt on the interpretation of high-grading estimates, and whether these potential levels of uncertainty are realistic.

5.3. Opportunities for future sampling

Production reports can be supplemented with data from the product grading system to provide the average weight (and therefore estimated number) of fish in the defined period. Data from individual fish passing over the grader are archived to generate a production report for any desired time period. Having direct observations of individual fish removes the need to infer an underlying size distribution, allowing for a direct comparison with size distributions of total catches by the Norwegian Reference Fleet. Chapter 6 of the Marine Resources Act 2008 gives a legal basis for a future implementation of a regulation that can allow for the acquisition of fish grading data from any fishing vessel for management or scientific purposes.

We have identified four issues regarding the use of individual fish measurements from grading machines. Firstly, there is the issue of cooperation with the fishing industry. We experienced some negative reactions from fishing companies to requests for aggregated data. Without addressing these issues, we would only expect concerns to increase if more detailed data were requested.

Secondly, the quantity of data generated by individual fish measurements would require a different solution for data transfer and storage. We have met with two leading companies that sell and maintain on-board fish grading systems to the Norwegian fishing industry. These companies can remotely access the system, and it is possible to transfer data across the connection. However, it is vital that agreements are developed between the Norwegian Directorate of Fisheries, fishing industry and grading technology companies to agree on a trusted and safe routine.

Thirdly, we need to develop a robust sampling design to ensure the reliability of estimates. For example, numerous studies in Norwegian offshore fisheries have stressed the importance of increasing the number of vessels (Helle and Pennington, 2004) and trips (Aanes and Pennington, 2003) sampled, demonstrating it is unnecessary to take large samples of fish from individual hauls (Pennington and Helle, 2011). A devoted pilot study would help to define the optimal sampling design, considering costs and excessive collection of sensitive data.

Finally, it is possible for fish to bypass the grading machine, and there is a risk that those fish bypassed may differ from those observed. We have no direct knowledge of when this occurs, but it is possible to infer by comparing grading data with production report summaries that include the bypassed fishes. Total weights in production reports are estimated, but if all fish were graded then the two totals should be equal within an acceptable degree of uncertainty.

Since 2019, the offshore segment of the Norwegian Reference Fleet began sampling discards, fishmeal and landed catches separately. High-grading could be identified by comparing the size distributions of these fractions, which could then be extrapolated to unsampled vessels to quantify the extent of high-grading in the fishery. Whilst this analysis could be done using only observations from the Norwegian Reference Fleet, we argue that individual size measurements from graders could supplement the analysis to include more vessels (Helle and Pennington, 2004).

Unfortunately, individual observations from the fish grader do not provide us with the ultimate need for length-based observations. However, they can be estimated using empirical length-weight relationships. A government report on the future of fisheries control (NOU, 2019) proposes an automatic documentation system where catches are registered at the earliest possible point after hauling. Such a system would provide direct observations of gross catches before any processing, removing the need estimate unreported catches, and therefore removing all the issues met in this report. For example, observations of fish before processing would remove the uncertainty surrounding the conversion from product weight to round weight. Likewise, laser measurement of length would also remove uncertainties surrounding a weight-length conversion. Any concerns would then be regarding the reliability of such a system, which would be dependent on enforcement strategy. However, a comparison with an equivalent, reliable source (e.g. gross catches recorded by the Norwegian Reference Fleet) could help to evaluate the reliability.

6. Acknowledgements

We are grateful for the cooperation of the fishing industry in both pilot studies. We also thank Vigdis Rong for the large task of standardising and compiling data. We finally thank David Hirst and Jon Helge Vølstad for discussions on the statistical analysis of the data.

7. 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.

Batsleer, J., Hamon, K. G., van Overzee, H. M. J., Rijnsdorp, A. D., and Poos, J. J. 2015. High-grading and over-quota discarding in mixed fisheries. Reviews in Fish Biology and Fisheries, 25: 715–736.

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

Blom, G. 2014. A Norwegian guide for controlling bodies' inspection of the application of conversion factors on board fishing vessels. Directorate of Fisheries, 64 p.

Breivik, O. N., Storvik, G., and Nedreaas, K. 2017. Latent Gaussian models to predict historical bycatch in commercial fishery. Fisheries Research, 185: 62–72.

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 (Accessed 10 June 2021).

Delignette-Muller, M. L., and Dutang, C. 2015. fitdistrplus: An R package for fitting distributions. Journal of Statistical Software, 64: 1–34.

Dickey-Collas, M., Pastoors, M. A., and van Keeken, O. A. 2007. Precisely wrong or vaguely right: simulations of noisy discard data and trends in fishing effort being included in the stock assessment of North Sea plaice. ICES Journal of Marine Science, 64: 1641–1649.

Dingsør, G. E. 2001. Estimation of discards in the commercial trawl fishery for cod. University of Bergen.

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

Gilman, E., Perez Roda, A., Huntington, T., Kennelly, S. J., Suuronen, P., Chaloupka, M., and Medley, P. A. H. 2020. Benchmarking global fisheries discards. Scientific Reports, 10: 1–8. Nature Publishing Group UK.

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.

Helle, K., and Pennington, M. 2004. Survey design considerations for estimating the length composition of the commercial catch of some deep-water species in the northeast Atlantic. Fisheries Research, 70: 55–60.

Hirst, D., Aanes, S., Storvik, G., Huseby, R. B., and Tvete, I. F. 2004. Estimating catch at age from market sampling data by using a Bayesian hierarchical model. Journal of the Royal Statistical Society: Series C (Applied Statistics), 53: 1–14.

Hoare, D., Graham, N., and Schon, P. J. 2011. The Irish Sea data-enhancement project: comparison of self-sampling and national data-collection programmes-results and experiences. ICES Journal of Marine Science, 68: 1778–1784.

Huston, M., and Deangelis, D. 1987. Size Bimodality in Monospecific Populations: A Critical Review of Potential Mechanisms. American Naturalist - AMER NATURALIST, 129.

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

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

McBride, M. M., and Fotland, Å. 1996. Estimation of unreported catch in a commercial trawl fishery. Journal of the Northwest Atlantic Fisheries Society, 18: 31–41.

Mello, L. G. S., and Rose, G. A. 2005. Seasonal cycles in weight and condition in Atlantic cod (*Gadus morhua* L.) in relation to fisheries. ICES Journal of Marine Science, 62: 1006–1015.

Methratta, E. T., and Link, J. S. 2007. Ontogenetic variation in habitat association for four groundfish species in the Gulf of Maine –Georges Bank region. Marine Ecology Progress Series, 338: 169–181.

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 (in Norwegian; Accessed 10 June 2021).

NOU. 2019. Framtidens fiskerikontroll. Departementenes sikkerhets- og serviceorganisasjon 2019:21 Oslo (in Norwegian).

Pálsson, Ó. K. 2003. A length-based analysis of haddock discards in Icelandic fisheries. Fisheries Research, 59: 437–446.

Pennington, M., and Helle, K. 2011. Evaluation of the design and efficiency of the Norwegian self-sampling purse-seine reference fleet. ICES Journal of Marine Science, 68: 1764–1768.

Pérez Roda, M. A., Gilman, E., Huntington, T., Kennelly, S. J., Suuronen, P., Chaloupka, M., and Medley, P. 2019. A third assessment of global marine fisheries discards. FAO Fisheries and Aquaculture Technical Paper No. 633. Rome, FAO. 78 pp.

Perretti, C. T., Deroba, J. J., and Legault, C. M. 2020. Simulation testing methods for estimating misreported catch in a state-space stock assessment model. ICES Journal of Marine Science, 77: 911–920.

Pitcher, T. J., Kalikoski, D., Pramod, G., and Short, K. 2009. Not honouring the code. Nature, 457: 658–659.

Punt, A. E., Smith, D. C., Tuck, G. N., and Methot, R. D. 2006. Including discard data in fisheries stock assessments: Two case studies from south-eastern Australia. Fisheries Research, 79: 239–250.

Reid, D. G., Calderwood, J., Afonso, P., Bourdaud, P., Fauconnet, L., González-Irusta, J. M., Mortensen, L. O., *et al.* 2019. The best way to reduce discards is by not catching them! *In* The European Landing Obligation: Reducing Discards in Complex, Multi-Species and Multi-Jurisdictional Fisheries, pp. 3–26. Springer, Cham.

Rizzi, S., Gampe, J., and Eilers, P. H. C. 2015. Efficient Estimation of Smooth Distributions from Coarsely Grouped Data. American Journal of Epidemiology, 182: 138–147.

Rizzi, S., Thinggaard, M., Engholm, G., Christensen, N., Johannesen, T. B., Vaupel, J. W., and Lindahl-Jacobsen, R. 2016. Comparison of non-parametric methods for ungrouping coarsely aggregated data. BMC Medical Research Methodology, 16: 1–12. BMC Medical Research Methodology.

Stenevik, E. K., Høines, Å., Kvamme, C., Otterå, H., Salthaug, A., and Svendsen, B. V. 2020. Fangstprøvelotteriet 2019 - Erfaringer og foreløpige resultat. Rapport fra Havforskningen nr. 2020-17 (in Norwegian).

Venables, W., and Ripley, B. 2002. Modern Applied Statistics with S. Springer, New York. https://www.stats.ox.ac.uk/pub/MASS4/ (Accessed 19 May 2021).

FISKERIDIREKTORATET

Telefon: 55 23 80 00
E-post: postmottak@fiskedir.no
Internett: www.fiskeridir.no

Livet i havet - vårt felles ansvar