Unravelling the scientific potential of high resolution fishery data

Schreiber Plet-Hansen, Kristian; Larsen, Erling; Mortensen, Lars O.; Nielsen, J. Rasmus; Ulrich, Clara

Published in:
Aquatic Living Resources

Link to article, DOI:
10.1051/alr/2018016

Publication date:
2018

Document Version
Peer-review version

Citation (APA):
Schreiber Plet-Hansen, K., Larsen, E., Mortensen, L. O., Nielsen, J. R., & Ulrich, C. (2018). Unravelling the scientific potential of high resolution fishery data. Aquatic Living Resources, 31, [24]. DOI: 10.1051/alr/2018016

Unravelling the scientific potential of high resolution fishery data

Kristian S. Plet-Hansen[1]*, Erling Larsen[1], Lars O. Mortensen[1], J. Rasmus Nielsen[1], Clara Ulrich[1]

[1]Technical University of Denmark, National Institute of Aquatic Resources (DTU Aqua), Kemitorvet, DK-2800 Kgs. Lyngby, Denmark.

*Corresponding author: tel.: +45 21485009; fax: +45 35883333; e-mail address: kspl@aqua.dtu.dk (K.S. Plet-Hansen).

Abstract

Fisheries science and fisheries management advice rely on both scientific and commercial data to estimate the distribution and abundance of marine species. These two data types differ, with scientific data having a broader geographical coverage but less intensity and time coverage compared to commercial data. Here we present a new type of commercial data with high resolution and coverage. To our knowledge, the dataset presented in this study has never been used for scientific purposes. While commercial datasets usually include the total weight by species on per haul basis, the new data also include the commercial size class for the species landed, recorded directly on a haul-by-haul basis. Thus, this dataset has the potential to provide knowledge on landed fish with as high spatio-temporal resolution as when coupling logbooks and sales slips but with the addition of detailed knowledge on the size distribution. Such information may otherwise be obtained through on-board observer programmes but unlike the observers' data, the dataset presented here is routinely collected on most of the trips of the vessels involved, which means that the coverage of the data for the individual vessel is larger than observers' data. Furthermore, the risk of changes in fishing behaviour due to the presence of an observer on-board is avoided. This paper describes the coverage and completeness of the dataset, and explores the reliability of the data available. We conclude that the main limitation is the small number of fishing vessels covered by the program, but that the data from those vessels are accurate, detailed, and relatively reliable.

Key words: Fisheries; Haul-by-haul information; Science-industry cooperation; Sea-packing commercial fishery data; Size distribution; Spatial and seasonal selectivity.

Introduction

Fisheries science and management rely on scientific survey data and commercial fishery data to estimate the status of marine populations and assess the impact of fishery on the environment. A key challenge is that the two data sources differ much in quality and detail. Scientific survey data usually have a broader and more homogeneous geographical coverage than commercial fishery data, as fishers target certain species and areas. However, scientific survey data have less intensity and temporal coverage (Pennino et al. 2016; Bourdaud et al. 2017). While both commercial and scientific data are important sources of information, it is a challenge to link the two types of data and provide a coherent picture (Poos et al. 2013; Bourdaud et al. 2017). Currently, integrated commercial datasets rely on coupling data from logbooks, sales slips and the Vessel Monitoring System (VMS) to allocate landings to vessels' hauls and fishing grounds (Hintzen et al. 2012). However, size composition at haul level is not known, and it is usually assumed that it is the same as the aggregated size composition from the entire trip (Bastardie et al. 2010). Fishing trips can cover several days and large areas, with potentially large variation in size composition; hence, these estimates probably introduce a bias. Thus, expanding the commercial data to incorporate accurate recordings of size at haul level could add significant quality to the information available (Verdoit et al. 2003; Bourdaud et al. 2017). A Danish initiative of packing-at-sea came to our attention that might be able to provide such information. The project started in 1995 with the purpose of investigating whether sea-packing could provide additional profit to fishers, by reducing their costs of size-sorting and packing at the auctions, and by ensuring higher quality fish. The project found a reduction in costs of 6-7% when packing fish at-sea but remained inconclusive on whether sea-packing resulted in a profit increase (Frederiksen and Olsen 1997; Frederiksen et al. 2002). Because sea-packed fish are labelled with information on size class, species, weight, vessel, and catch time, a by-product of this project was the development of a database collecting the size composition of landings at the haul level together with detailed spatio-temporal information. Although on-board observers programmes in the EU collect data with similar resolution and characteristics, the sea-packing data extends the data coverage substantially because vessels engaged in sea-packing record their sea-packed landings for most trips, while observers only record a limited number of trips. Additionally, sea-packing data are collected by fishers, without additional costs to be borne by scientists or public authorities.

In 2002, the Council of the European Union laid down rules for increased traceability of food goods, including fish (EU 2002). The traceability regulations apply for batches of fish, with a batch being a quantity of fish caught at one time. The regulations do allow for the registration of a batch as the compiled landings from a full fishing trip. Additionally, spatial traceability regulations are complied with if a batch can be traced to the fishing area (e.g. an ICES subdivision) which covers large areas. In Denmark three traceability systems were developed to meet the requirements; the Vessels Data Exchange Center (VDEC) software, the yellow catch information notes and the "Sporbarhed i Fiskerisektoren" (SIF) database, which is an add-on to the sea-packing project. The VDEC is in theory capable of delivering more detailed data than the electronic logbook (eLog), including crate landing composition and size classes (a crate is a standard size box used to store fish for landing (Pack and Sea A/S 2018)). However, in practice, most of the data reported in the VDEC are limited to haul position, time, and non-sized landings information (O. Skov, personal communication). The yellow catch information notes were developed by the industry to ensure compliance with the regulations among vessels unfit for carrying sea-packing or VDEC equipment (Dandanell and Vejrup 2013). A note is filled in for the crate with information of the fishing trip including date of first and last fishing, geographical area where fishing took place (as ICES subdivision), gear type and other administrative information, as well as the species and commercial size class. The minimum labelling and information requirements are thus complied with (EU 2001, 2009, 2011; Dandanell and Vejrup 2013).

The present study focus on the third system, the SIF database. We analyse and explore the accessibility, coverage, consistency and reliability of the data, in order to assess whether it may be used for scientific studies and in management advice. The quality of the data is assessed by comparing it with the eLog, sales slips and data from a trial using Remote Electronic Monitoring with a CCTV camera system (EM). The objective of the present paper is only to investigate whether SIF data are suitable and reliable, before they can be used in future studies. As such, we primarily focus here on describing these new data and assess their quality. Future studies involving SIF data are briefly suggested, including comparison with coupled VMS and logbook data as well as studies on spatial size distribution for certain species.

Materials and Methods

The SIF database

The SIF database began in 2012 as collaboration between the Danish Fishermen's Association (DFPO), the Danish AgriFish Agency and the retail industry. The sea-packing data in SIF provide information at haul level on the landed species and size composition by weight, together with detailed information on date, time and position of the haul. The size classes applied are those defined by the EU regulation and size classes used by the fish auctions (Table 1) (EU 1996; Danske Fiskeauktioner 2017). The sea-packing equipment includes a dynamic scale, which records the weight of each size class of each species automatically. When in port, the records are relayed online from the sea-packing software to SIF. The weight recorded by the sea-packing equipment is the gutted weight, not the live weight as recorded in the eLog (Frederiksen et al. 1997, 2002; Danish AgriFish Agency 2017). As in the eLog, the SIF database allows for entries of discards in addition to the landings. Figure 1 presents a schematic of the difference between landings information at haul level in the eLog and SIF. SIF provides the size composition of the landings directly at haul level, assuming that the sea-packed fish of a given species are representative of the total landings of that species in the individual haul. This assumption will be discussed in the subsection *Using SIF data*. SIF is linked with the eLog, from which the temporal and spatial data for the hauls are derived. In 2016, funding for SIF operational costs was reduced. The future of SIF is thus uncertain, although it recently proved valuable. In 2017, the German authorities required traceability data for a batch of fish a German buyer had purchased from a wholesaler in Denmark. The required information could be retrieved from in SIF and met the expectations of the German authorities, thus demonstrating the operationality of the system (C. S. Pedersen, personal communication).

Data collection

As each vessel owns its own data in SIF, individual acceptance to use the data for the present study was required. Around 90 vessels operated with sea-packing in Denmark in 2015 and 2016. All sea-packing vessels were part of the large-scale fleet, which consisted of 419 vessels in 2015 and 396 vessels in 2016 (STECF 2017). However, due to confidentiality agreements, vessel details from SIF could not be provided by the database administrator (C. S. Pedersen, personal communication). 28 vessel owners have thus been personally contacted so far, and asked whether they sea-pack their landings and are willing to grant access to their SIF data.. At the time of writing, confirmation was still pending from four skippers, 13 skippers had granted access to their SIF data and 11 skippers had refused (Table 2). The access to SIF occur through a website, with no export function. A web scraper was thus developed to extract the data.

Study period

The study period is January 1 2015 to December 31 2016. Over this period, high resolution haul data for five vessels and SIF data could be compared with electronic monitoring (EM) data (GPS) for two vessels, which both had sea-packing equipment and participated in the Danish Cod Catch Quota Management trial (Ulrich et al. 2015; Bergsson and Plet-Hansen 2016; Bergsson et al. 2017).

Assessing validity of SIF against DFAD and eLog

For the validity assessment, SIF data from vessels A, B, C, D and E in 2015 and 2016 were compared to the DTU AQUA DFAD (Danish Fisheries Analyses Database) dataset. DFAD is based on sales slips merged with the eLog catches and fleet register data. Catches are recorded as total live weight of each species and since 2015 it has been mandatory to record catches in the eLog on a haul-by-haul level (EU 2011; Danish AgriFish Agency 2017). The coupling of eLog haul data and sales slips data do allow for inference of landings' size composition at the haul level assuming constant size distribution across all hauls (Bastardie et al. 2010; Hintzen et al. 2012). However, the assumption of even size distribution risks assigning inaccurate size distributions to the haul.

Not all species landed by a vessel are sea-packed. To analyse the completeness of the SIF data the species recorded in SIF were compared to the same vessels' data from DFAD. The 10 most important species (in landings by weight) for the five vessels were identified based on DFAD landings records. These 10 species constituted 95.8% of the landings by weight for the five vessels in both years. The completeness of landings recorded in SIF compared to DFAD was calculated as:

$$C_L = 100 - \frac{L_{DFAD} - L_{SIF}}{L_{DFAD}} * 100 \quad (1)$$

Where L is the sum of recorded landings of the species in DFAD and SIF respectively. No conversion factor was needed for the comparison, since both SIF and DFAD have records of the gutted weight.

Similarly, the completeness of hauls available in SIF was estimated based on the number of hauls according to the eLog, using:

$$C_H = 100 - \frac{H_{eLog} - H_{SIF}}{H_{eLog}} * 100 \quad (2)$$

Where H is the number of recorded hauls in eLog and SIF respectively.

A comparison between SIF and DFAD of the species and commercial size classes recorded by vessel A, B, C, D and E during 2015 and 2016 for the 10 most landed species was then performed. SIF and DFAD data were merged based on the trips' landing date. The weight of each commercial size class of the 10 most landed species for each trip was summed based on the unique logbook number identifying each fishing trip. Trips with no records in either SIF or DFAD were excluded. The largest size class for cod (*Gadus morhua*) and hake (*Merluccius merluccius*) in SIF is 0, whereas the largest size class is 1 in DFAD (Table 1). The division between the second largest size class, size class 2, and size class 1 is the same for SIF and DFAD. Therefore, size class 0 was aggregated with size class 1 in SIF to render the comparison between databases possible. In addition to a visual comparison of SIF and DFAD data at trip level, the fit between SIF and DFAD records was analysed with a linear model using the *lm* function in R. This was done to estimate how close SIF records are to DFAD records and vice-versa. A log-transformation was applied to landings recorded in SIF and DFAD whereby normal distribution was induced.

The model is thus written as:

$$\log(y_i) = a + \log(x_i) * b \tag{3}$$

Where a is the intercept, b is the slope, *y* is the landings by size class recorded in SIF, *x* is the landings by size class recorded in DFAD and *i* is an index for the fishing trip and commercial size class of the investigated species.

Essentially, DFAD should contain all landings of all species from all the vessels' fishing trips. SIF has only records of all landings of all species from when the vessel started sea-packing during the fishing trip. A comparison of the trip-based percentwise size class compositions of landings was performed between trips where sea-packing did not take place and trips where sea-packing was conducted. This was done to investigate whether a potential bias in the size class compositions is possible depending on whether a vessel packs at-sea or not. The comparison was made solely using DFAD, because SIF does not have information in trips without sea-packing. First, the size class composition of the landings recorded in DFAD was calculated as a percentage of the total landings recorded in DFAD for trips where SIF records also existed and for trips where SIF records did not exist. This was plotted and investigated visually. Then, a non-parametric analysis was performed using the Wilcoxon rank-sum test, to detect potential bias in size distribution which could occur if fishers for instance only sea-pack at trips with ample volumes of large fish.

To investigate the effect of year, vessel and size class on the differences between landings recorded in SIF compared to DFAD, an extension of the model in equation 3 was made and analysed using an Analysis of Covariance (ANCOVA). The model is written as:

$$\log(y_i) = \log(x_i) + \beta_1(\mu_i) + \beta_2(\nu_i) + \beta_3(s_i) \quad (4)$$

Where *y* is the landings by size class recorded in SIF, *x* is the landings by size class recorded in DFAD, *i* is an index for the fishing trip, μ is year, ν is vessel, s is size class and β_1 to β_3 are the effects of year, vessel and size class for the investigated species.

Spatial distribution of SIF data compared to EM data

Because the SIF system depend on the eLog for the temporal and spatial haul information, a geographic comparison with DFAD is not relevant. Therefore, coverage quality was assessed using a different dataset, comparing SIF with the GPS sensor data from an EM trial run by the Danish AgriFish Agency in 2015 and 2016 (Bergsson and Plet-Hansen 2016; Bergsson et al. 2017). This was done for two vessels that took part in this trial during 2015 and 2016. EM GPS data were plotted as dots at a 1-minute interval. Start and end position according to SIF was used to plot lines for each haul on the same chart. Because this assumes linear track courses, some deviance is expected. Additionally, some hauls with unrealistic haul lengths and towing speeds were spotted in SIF. SIF hauls were excluded if the towing speed exceeded 7 knots. The criteria for exclusion was based on information from the vessel owners on their maximum and usual towing lengths as well as an inspection of the maximum towing speeds recorded in the EM trial. In addition to the visual inspection, the mean mid-latitude and mid-longitude were calculated for each haul. Because fishers target certain fishing grounds, the distribution of fishing hauls becomes non-random and it is not possible to induce normal distribution of samples. Therefore, statistical comparison of mid-latitude and mid-longitude was performed using a Wilcoxon rank-sum test.

Results

Although it is possible to enter discards in SIF, none of the investigated vessels had any discards recorded. Seven of the 13 skippers who granted access to their SIF data had recordings at the haul level with high resolution, while the data from the other six showed that on these vessels, the sea-packing equipment was not used in a manner where the size classes were recorded at the haul level. The main reason given for this was that the vessels had used the sea-packing equipment to clean the fish during their catch processing but had not stored their landings in size-graded crates (Table 2). This was also the main reason given by the 11 skippers who have not granted access.

Species not occurring in SIF

Of all species reported in DFAD for each vessel, only a few were never reported in SIF. For vessel A, this was the case for five species: Atlantic mackerel (*Scomber scombrus*), edible crab (*Cancer pagurus*), marine crabs (*Brachyura sp.*), greater weever (*Trachimus draco*) and lumpfish (*Cyclopterus lumpus*). For vessel B six species: Norway lobster (*Nephrops norvegicus*), golden redfish (*Sebastes marinus*), greater forkbeard (*Phycis blennoides*), long-rough dab (*Hippoglossoides platessoides*), cuttlefish (*Sepiidae sp*.) and tope shark (*Galeorhinus galeus*). For vessel C and D three species: Atlantic mackerel, edible crab and lumpfish. For vessel E five species: Norway lobster, golden redfish, lumpfish, greater forkbeard and blue ling (*Molva dypterygia*). The weight of the species never recorded in SIF ranged from 0.02% (vessels C and E) to 0.1% (vessel B).

Comparison of trips, hauls and 10 most landed species

The majority of hauls and trips were represented in both SIF and DFAD, although a third of the 14,570 species*haul combinations were missing in SIF (Table 3). For the reported landings, the highest completeness C_L was achieved for vessel B at around 90% on average, followed by vessel A at around 80% on average, whereas vessel C had the poorest completeness, at 69%. Overall the size class composition was similar on an aggregated level (Figure 2) but the means differed significantly in 16 out of 39 cases when $\alpha = 0.05$ (Table 4). For cod, hake, haddock (*Melanogrammus aeglefinus*), lemon sole (*Microstomus kitt*), turbot (*Scophthalmus maximus*) and witch flounder (*Glyptocephalus cynoglossus*), the size classes constituted roughly the same percentage of the overall landings regardless of whether the trips had only DFAD data or had SIF too. The largest overall discrepancy was for saithe (*Pollachius virens*) where size class 3 constituted a lower percentage of the landed weight while size class 4 constituted a larger share when trips had not been recorded in SIF. However, all species had at least one size class with a significant difference in percentwise composition. Conversely, all species also had at least one size class where no significant difference was found. Additionally, the standard deviation was large for all species and size classes, meaning that large variation in size composition occur between trips.

Log-transformation of landings recorded in SIF and DFAD was necessary to assume normal distribution (Figure 3). A scatterplot and a linear model fit was made for the size classes of the 10 investigated species of each vessel at trip level (Figure 4 and Table 5). Saithe, turbot, witch flounder, wolffish (*Anarchichas spp.*) and monkfish (*Lophius spp.*) had R^2-values and a scatterplot close to a 1:1 ratio between SIF and DFAD by trip for most vessels. However, monkfish was not sorted into size classes on vessel A when sea-packed, and vessel E had several trips with a poor fit for the medium size classes of saithe as well as some trips with a poor fit for the largest size class of wolffish. Correlations were also generally high for hake and lemon sole but vessel D had several trips where the larger size classes of these two species had a poor fit. Vessel A also had some trips with a poor fit for lemon sole, and this species was rarely landed for vessel B. Haddock had high R^2-values as well but not for all years and all vessels, where especially vessel B and D in 2016 had a poor fit. Cod had R^2-values and a scatterplot with a good agreement between SIF and DFAD for vessel B, but not for the rest of the vessels. For plaice (*Pleuronectes platessa*) the scatterplot and R^2-values were poor for most vessels. Interestingly, some occurrences of more landings in weight in SIF than DFAD appeared, mainly for witch flounder, which in theory should not be possible, since the summing of all SIF data should also be found in the total recorded landings for any given trip. Presenting this to the fishers revealed two reasons; 1) small mismatches are inevitable, as the fishery auctions, from where the landings data in DFAD are derived, only record landings in total kilograms, whereas the sea-packing equipment uses scales with dynamic motion compensation and relay data with two decimals. 2) Larger mismatches could be an artefact in the SIF system. If a crate is labelled wrongfully, e.g. by recording the wrong size class or species, a new label must be made. This in turn will be recorded as a new entry in SIF and the fishers cannot delete the old entry, meaning that the same crate will count twice in SIF.

Extension of the model to include the effect of year, vessel and size class revealed that each of these factors could have a significant effect among the species (table 6). The effect of year was significant for cod, hake and lemon sole. Vessel effect was significant for all species, except haddock and turbot and the effect of size class was significant for all species, except witch flounder. The log-transformed landings in DFAD had a significant effect and the largest sum of squares and F-value for all species.

Spatial distribution of hauls compared to EM data

The exclusion criteria to filter for unrealistic haul lengths and towing speeds in SIF led to the exclusion of respectively 91 and 71 hauls for the two EM vessels, corresponding to 6.33% and 7.67% of recorded hauls. Overlay maps for positions according to EM GPS data and according to SIF in 2015 and 2016 are presented in Figure 5. Visually, most areas had overlap between SIF and EM but in 2015, the difference between positional data in SIF and EM was statistically significant (table 7). An area at roughly 59° N and 0.5° W was visually identified where fishing took place according to EM but no hauls have been recorded in SIF, neither in 2015 nor in 2016.

Discussion

The SIF dataset possess information not available in the currently used commercial fisheries data. That cover direct observations on size distributions at the haul level instead of merely at the trip level. The completeness of SIF compared to DFAD shows overall a good match, albeit not perfect. Although all five vessels landed a few species that were never sea-packed and, consequently, present in DFAD but not in SIF, these species only constituted a minor fraction of the vessels' total landings. Thus, they were non-target species for the vessels. According to the fishers, vessels engaged in sea-packing may choose not to sea-pack a species if it is not considered worth the effort of sea-packing during the catch processing. Norway lobster is an example of a potential target species that is not necessarily sea-packed. This is because as the added value is not considered to be large enough, which is also the case for several flatfish species.

Fishing trips and hauls recorded in the eLog were overall well represented in SIF. No discards were recorded in SIF, which is likely because the legal purpose of the dataset is for traceability requirements of the landings.

Several trips had records of landings for one or more of the 10 investigated species in DFAD but no records of the species in SIF. A reason for this may be the loss of data when merging DFAD and SIF, because there are no unique haul and trip IDs shared between SIF and DFAD. Therefore, the common identifier used to merge SIF and DFAD was the landings date, which can be inferred from SIF and is recorded in the DFAD data. Mismatch may also be due to lack of vessel storage capacity to pack all their landings in crates at-sea. Because it takes up more storage room to sea-pack landings there is a trade-off between continuing to fish after the storage capacity for sea-packing is reached. On the one hand, sea-packing should give a higher quality and thereby higher price for the landings (Frederiksen and Olsen 1997; Frederiksen et al. 2002). On the other hand, the cost of steaming between fishing grounds and port may make it more profitable to continue fishing, store landings in larger bulks, and land a larger amount of unsorted fish, which will give a higher total revenue. The choice between one and the other is likely to be influenced by several factors. These include among others as the amount of remaining quota, the expected value of the landings already in storage, how far into the expected duration of the fishing trip a haul takes place, and the weather conditions. Accordingly, there may not necessarily be consistency between fishing trips as to whether a species is sea-packed or not. The fact that plaice is the species where SIF records are poorest supports this, as plaice is a relatively low value species in this context. Conversely, it is likely that species with a high profit gained from sea-packing will have the best agreement between DFAD and SIF records. Monkfish has good agreement for most vessels, which supports the above perspective as monkfish has a relatively high value. The model extension to include the effect of year, vessel, and size class for each species did not reveal which factors specifically and significantly influence the choice of sea-packing or not. The model output show that factors other than year, vessel, and size class significantly influence the lack of a perfect fit between SIF and DFAD records. As stated above, external factors may well heavily influence the choice. This include factors that may vary substantially such as fish price. Furthermore, due to the Danish Individual Vessel Quota system, it is difficult to specify the remaining quota during a year, which may also influence the choice. We, nonetheless, consider it to be beyond the scope of this study to further analyse these factors here. Future studies on the frequency of storage limitations, possible correlation between expected fish prices and sea-packing, or cost-benefit analysis of the added workload at-sea compared to the potential gain from sea-packing could shed further light on the underlying reasons and key driving factors behind the frequency of trips with landings recorded in DFAD while lacking in SIF. The potential bias created by lack of SIF records for certain trips seems limited, though. Overall, there are only small differences in the percentwise size composition in the landings for the DFAD dataset when looking at trips where SIF data was available compared to trips with no SIF data available. However, statistical test output of the percentwise composition suggest large variation among trips. As a whole, the investigations and tests comparing SIF and DFAD revealed that a consistent bias in SIF records seems unlikely. Lack of entries in SIF varies between vessels, years, species and possibly size classes, although fishers have stated that they either do not sea-pack a species or sea-pack all retained specimens at the hauls where they sea-pack. In light of this, SIF should not be viewed as a full record but rather as a subsample of the landings with higher resolution for certain species. Due to the species-to-species variation in reliability in SIF, studies utilizing SIF data should verify the completeness of the specific SIF data available for those species, which are to be investigated, prior to any further analysis.

Spatial data

Overall, there is a good spatial overlap between the SIF and EM datasets. However, some gaps in spatial coverage occur, and a statistically significant difference between mid-points of hauls was found for 2015. Several reasons can explain the discrepancy. First, hauls recorded in SIF with unrealistic duration and towing speeds were excluded which inevitably creates gaps for SIF compared to EM. Second, positional data in SIF is exported from the eLog. Although the eLog software allow for real-time entries of the vessel's position, the skipper may postpone entries of haul data, including fishing time and position, as long as the data has been entered prior to the mandatory deadline of data transmission (once every 24 hours). Therefore, a certain mismatch could be caused by human errors if positional data is entered manually in the eLog. Third, there is an inherent error in plotting a haul as a simple straight line from haul start to end. Adjustments in vessels' course and drag will mean that towing paths are not conducted in straight lines in the real world, which can cause mismatch when assuming a straight line between start and end position of the haul. Fourth, some gaps may come from fishers testing an area for fish. If the catch in this area is poor, then no sea-packing will occur, meaning that no haul is recorded in SIF, but because a fishing activity was recorded in EM, the haul will appear in the EM data. This could explain the mismatch in an area around 59° N and 0.5° W. Fifth, the spatial resolution of the data used for the statistical test will influence the outcome of the test of means. Finally, breakdowns have happened in the GPS equipment during the EM trial, meaning that it is possible for hauls to have taken place and be present in SIF without being recorded in EM.

Using SIF data

When taking the differences in data between DFAD and SIF into account, it is clear that the quality of the SIF data has to be scrutinized at the vessel and species level before it can be utilized for scientific and management purposes. Spatial and temporal entries in SIF seem valid, but due to inaccurate reporting, it is necessary to filter out hauls where spatial or temporal records are unrealistic. This can be done by setting up exclusion criteria and filtering by these. Prior to in depth analysis of species distributions it is necessary to validate the species records in SIF for the individual year, vessel, species, and size class. The agreement between DFAD and SIF can vary substantially. The discrepancies originating from incorrect crate labelling are more difficult to remove. It is a very species and vessel specific issue and therefore only relate to analysis for these specific species, e.g. witch flounder. The simplest approach is to exclude the records from the problematic vessel and/or species, depending on the analysis. The more cumbersome solution is to identify the trips where incorrect labelling has happened, as can be done for the trips where SIF do not contain the majority of landings of a species. By identifying the vessel, species and size class, one can find the corresponding landings in DFAD and SIF and subset for these. Then, using the landings date, the corresponding hauls for the specific fishing trip can be removed from the dataset.

Based on talks with sea-packing fishers, species are generally either sea-packed at the haul level or not at all. Mismatch between SIF and DFAD at the trip level should be due to hauls where species where not sea-packed rather than hauls where a fraction of a species was sea-packed. However, the effect of size class in the extended model does not fully support this statement.

Possible applications

There are clear limitations regarding the usefulness of SIF owing to the facts that i) the future of SIF is uncertain due to funding issues, ii) the majority of Danish fishing vessels do not use it, and iii) vessels can refuse to share SIF data. Furthermore, several vessels with sea-packing do not complete the entries into SIF in a manner that allow for better spatial resolution than DFAD. The relatively short time coverage of SIF further limits its use. Nevertheless, SIF have several benefits: SIF data is already collected and is therefore a free data source, which only requires the time spent on access permission and adjustment of a web scraper to collect. SIF does not serve as a direct control measure but is used for commercial purposes and to fulfil traceability requirements, whereby there should be little if any incentive to tamper with the system. This study serves, therefore, as a proof of concept that it is possible to obtain precise size distribution from fisheries data at the haul level, even though it is not a legal requirement. Indeed, the fisheries control in Greenland already requires vessels above 75 GRT to include the size distribution of the landings at the haul level (Greenland's Autonomy 2010). Although the number of sea-packing vessels is low in Denmark, the landed volume from sea-packing vessels is large and the activity coverage is extensive. The five Danish vessels investigated in this study have SIF data from 258 trips in 2015 and 293 trips in 2016. In 2015 and 2016, the entire Danish observer programme covered a total of 224 and 262 trips respectively. When SIF and observer data overlap, SIF could also be used to investigate potential behavioural aspects of observer presence. Because fishers may refuse to take observers on-board, there is a risk of a bias in the observer data relative to the reason for not wanting observers. Likewise, fishers may adapt their fishing behaviour while carrying observers, either intentionally or unintentionally, which may also cause a bias in observer data. While sharing SIF data with scientist or fisheries managers is purely voluntary, there is an economic incentive to conduct sea-packing as costs are reduced (Frederiksen and Olsen 1997; Frederiksen et al. 2002) and vessels are liable to the fish auctions for correct labelling of sea-packed landings. Therefore, the risk of fishers adapting fishing behaviour is less likely for SIF. Investigations with SIF data could enhance the knowledge on spatial explicit fish distributions, for instance by mapping areas with a larger share of juveniles for certain species, whereby fishers may improve their spatial selectivity. Especially monkfish and wolffish could be of interest for analysis utilizing the size class information in SIF as these species are data poor and have some of the best records in SIF for the investigated species.

Based on the presented results, the next planned step in utilization of SIF data is to compare the spatial and temporal distribution of size classes for species well represented in SIF data, to that of DFAD and VMS-logbook coupled data. This will allow for testing the validity of the homogeneous reallocation of size classes, as well as showing the importance of having the size composition at the haul level.

Conclusion

SIF provides new, relatively reliable data on the size composition of important commercial fish species with the same or higher resolution than what is available in traditional fisheries data. However, the quantity, quality and reliability vary between vessels and species. Although SIF has high coverage and detailed landings and spatio-temporal information, the dataset has limited coverage in the number of vessels. If the SIF database is maintained and SIF data continuously collected, we believe SIF could provide additional knowledge on detailed spatial patterns of fishing effort and commercial fish species and size distributions. Because SIF provide direct observations at the haul level it could be used for analysis at a vessel or métier specific level, for instance on catchability, spatial selectivity, seasonal patterns or to compare and verify outcomes of spatial fishery evaluation models as evaluated in Nielsen et al. (2018). A fleet-wide application or stock assessment usage would require an expansion of the vessel coverage and better accessibility to SIF data. It is our hope that this study may serve as a case study to highlight the possibilities that exist in enhancement of commercial fisheries data available to science.

Acknowledgements

This work has received funding from the Horizon 2020 Programme under grant agreement DiscardLess number 633680 and from the EASME project DRuMFISH contract number EASME/EMFF/2014/1.3.2.4/ SI2.721116. This support is gratefully acknowledged. The authors thankfully acknowledge the support from fishers who have been willing to share their data as well as the parties who have made such data sharing possible. The help and information provided by Lyngsoe Systems, Anchor Lab K/S, DFPO and the Danish AgriFish Agency is gratefully acknowledged. Carsten Søndergaard Pedersen from Pack and Sea A/S and Kim Clemmensen deserve special thanks for their continuous support.

References

Bastardie F., Nielsen J.R., Ulrich C., Egekvist J., Degel H., 2010, Detailed mapping of fishing effort and landings by coupling fishing logbooks with satellite-recorded vessel geo-location. Fish. Res. 106, 41–53.

Bergsson H., Plet-Hansen K.S., 2016, Final Report on Development and usage of Electronic Monitoring Systems as a measure to monitor compliance with the Landing Obligation - 2015.

Bergsson H., Plet-Hansen K.S., Jessen L.N., Jensen P., Bahlke S.Ø., 2017, Final Report on Development and usage of REM systems along with electronic data transfer as a measure to monitor compliance with the Landing Obligation – 2016.

Bourdaud P., Travers-trolet M., Vermard Y., Cormon X., Marchal P., 2017, Inferring the annual , seasonal , and spatial distributions of marine species from complementary research and commercial vessels ’ catch rates. ICES J. Mar. Sci. 74, 2415–2426.

Dandanell R., Vejrup K., 2013, TEMA om sporbarhed i fiskeriet. Fisk. Tid. Available online at: https://issuu.com/dandanell/docs/sporbarhed_i_fiskeriet_tema_i_fiskeri_tidende_feb. (accessed 1.10.18).

Danish AgriFish Agency, 2017, Elektronisk logbog. Available online at: http://lfst.dk/fiskeri/erhvervsfiskeri/indberetning-og-foering-af-logbog/elektronisk-logbog/ (accessed 12.21.17).

Danske Fiskeauktioner, 2017, Grading. Available online at: http://www.dfa.as/sortering (accessed 12.21.17).

EU, 1996, COUNCIL REGULATION (EC) No 2406/96 of 26 November 1996 laying down common marketing standards for certain fishery products. Off. J. Eur. Communities L334/1, 1–15.

EU, 2001, COMMISSION REGULATION (EC) No 2065/2001 of 22 October 2001 laying down detailed rules for the application of Council Regulation (EC) No 104/2000 as regards informing consumers about fishery and aquaculture products. Off. J. Eur. Union L 278/6.

EU, 2002, REGULATION (EC) No 178/2002 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 28 January 2002 laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food saf. Off. J. Eur. Communities L31/1.

EU, 2009, COUNCIL REGULATION (EC) No 1224/2009 of 20 November 2009 establishing a Community control system for ensuring compliance with the rules of the common fisheries policy, amending Regulations (EC) No 847/96, (EC) No 2371/2002, (EC) No 811/2004, (EC) No 768/2. Off. J. Eur. Union L 343/1.

EU, 2011, COMMISSION IMPLEMENTING REGULATION (EU) No 404/2011 of 8 April 2011 laying down detailed rules for the implementation of Council Regulation (EC) No 1224/2009 establishing a Community control system for ensuring compliance with the rules of the

Common Fish. Off. J. Eur. Union L 112/1.

Frederiksen M., Osterberg C., Silberg S., Larsen E., Bremner A., 2002, Info-Fisk. Development and Validation of an Internet Based Traceability System in a Danish Domestic Fresh Fish Chain. J. Aquat. Food Prod. Technol. 11, 13–34.

Frederiksen M.T., Olsen K.B., 1997, Søpakning med sporbar deklaration, Lyngby: Danmarks Fiskeriundersøgelser. (DFU-rapport; Nr. 45-97).

Frederiksen M.T., Olsen K.B., Popescu V., 1997, Integrated Quality Assurance of Chilled Food Fish at Sea. In: Luten J.B., Børresen T., Oehlenschläger J. (Eds.), Seafood From Producer To Consumer, Integrated Approach To Quality. Amsterdam, Elsevier, pp. 87–96.

Greenland's Autonomy, 2010, Selvstyrets bekendtgørelse nr. 18 af 9. december 2010 om kontrol med havgående fiskeri. Available online at: http://lovgivning.gl/lov?rid=%7B8EC0C382-4157-4543-AEE1-E83AB40CABEE%7D (accessed 12.21.17).

Hintzen N.T., Bastardie F., Beare D., Piet G., Ulrich C., Deporte N., Egekvist J., Degel H., 2012, VMStools: Open source software for the processing, analysis and visualization of fisheries logbook and VMS data. Fish. Res. 115–116, 31–43.

Nielsen J.R., Thunberg E., Holland D.S., Schmidt J.O., Fulton E.A., Bastardie F., Punt A.E., Allen I., Bartelings H., Bertignac M., Bethke E., Bossier S., Buckworth R., Carpenter G., Christensen A., Christensen V., Da-Rocha J.M., Deng R., Dichmont C., Doering R., Esteban A., Fernandes J.A., Frost H., Garcia D., Gasche L., Gascuel D., Gourguet S., Groeneveld R.A., Guillén J., Guyader O., Hamon K.G., Hoff A., Horbowy J., Hutton T., Lehuta S., Little L.R., Lleonart J., Macher C., Mackinson S., Mahevas S., Marchal P., Mato-Amboage R., Mapstone B., Maynou F., Merzéréaud M., Palacz A., Pascoe S., Paulrud A., Plaganyi E., Prellezo R., van Putten E.I., Quaas M., Ravn-Jonsen L., Sanchez S., Simons S., Thébaud O., Tomczak M.T., Ulrich C., van Dijk D., Vermard Y., Voss R., Waldo S., 2018, Integrated ecological–economic fisheries models — Evaluation, review and challenges for implementation. Fish Fish. 19, 1–29.

Pack and Sea A/S, 2018, PACK AND SEA - Types of crates/tubs. Available online at: http://packandsea.dk/ (accessed 1.10.18).

Pennino M.G., Conesa D., López-Quílez A., Munoz F., Fernández A., Bellido J.M., 2016, Fishery-dependent and -independent data lead to consistent estimations of essential habitats. ICES J. Mar. Sci. 73, 2302–2310.

Poos J.J., Aarts G., Vandemaele S., Willems W., Bolle L.J., Helmond A.T.M. Van, 2013, Estimating spatial and temporal variability of juvenile North Sea plaice from opportunistic data. J. Sea Res. 75, 118–128.

STECF, 2017, Scientific, Technical and Economic Committee for Fisheries (STECF) – The 2017 Annual Economic Report on the EU Fishing Fleet (STECF-17-12).

Ulrich C., Olesen H.J., Bergsson H., Egekvist J., Håkansson K.B., Dalskov J., Kindt-larsen L., Storr-paulsen M., 2015, Discarding of cod in the Danish Fully Documented Fisheries trials. ICES J. Mar. Sci.

Verdoit M., Pelletier D., Bellail R., 2003, Are commercial logbook and scientific CPUE data useful for characterizing the spatial and seasonal distribution of exploited populations? The case of the Celtic Sea whiting. Aquat. Living Resour. 16, 467–485.

Figure captions

Figure 1. Conceptual figure of the difference between landings data available at haul level in the electronic logbook and the sea-packing data available in the SIF database.

Figure 2. Landings' size composition in percent stratified on trips with only DFAD data and trips with both DFAD and SIF data. Size class 1 are the largest specimens.

Figure 3. QQ-plot for I) log-transformed landings recorded in SIF. II) log-transformed landings recorded in DFAD.

Figure 4. Landings per trip according to DFAD and SIF for the 10 most landed species in 2015 and 2016 by species and commercial size class. Points: The aggregated weight of the species and size class for a fishing trip. The x-axis represent the weight according to DFAD, the y-axis represent the weight according to SIF. Blue dashed line: linear model fit between DFAD and SIF. Black line: The 1:1 ratio between DFAD and SIF. Size class 9 is unsorted.

Figure 5. Fishing activity overlap between EM and SIF for two vessels. I) 2015. II) 2016. Blue points: Fishing activity recorded by EM GPS sensors (1-minute interval). Yellow lines: Hauls according to SIF. The EM trial did not cover the Baltic Sea and the maps do therefore not include hauls in this area.

Table 1. Commercial fish size classes and their corresponding weight in kg for the 10 investigated species based on SIF and Danish fish auction as well as DFAD and EU regulations.

Species	**Size class, SIF/Danish fish auction**	**Weight range [kg/fish]**	**Size class, DFAD/EU regulation**	**Weight range [kg/fish]**
Cod	0	>10.00		
	1	7.00-10.00	1	>7.00
	2	4.00-7.00	2	4.00-7.00
	3	2.00-4.00	3	2.00-4.00
	4	1.00-2.00	4	1.00-2.00
	5	0.30-1.00	5	0.30-1.00
Hake	0	>4.00		
	1	2.50-4.00	1	>2.50
	2	1.20-2.50	2	1.20-2.50
	3	0.60-1.20	3	0.60-1.20
	4	0.28-0.60	4	0.28-0.60
Plaice	1	>0.60	1	>0.60
	2	0.40-0.60	2	0.40-0.60
	3	0.30-0.40	3	0.30-0.40
	4	0.15-0.30	4	0.15-0.30
Haddock	1	>1.00	1	>1.00
	2	0.57-1.00	2	0.57-1.00
	3	0.37-0.57	3	0.37-0.57
	4	0.17-0.37	4	0.17-0.37
Saithe	1	>5.00	1	>5.00
	2	3.00-5.00	2	3.00-5.00
	3	1.50-3.00	3	1.50-3.00
	4	0.30-1.50	4	0.30-1.50
Lemon sole	1	>0.60	1	>0.60
	2	0.35-0.60	2	0.35-0.60
	3	0.18-0.35	3	0.18-0.35
Monkfish	1	>8.00	1	>8.00
	2	4.00-8.00	2	4.00-8.00
	3	2.00-4.00	3	2.00-4.00
	4	1.00-2.00	4	1.00-2.00
	5	0.50-1.00	5	0.50-1.00
Turbot	1	>3.00	1	>3.00
	2	2.00-3.00	2	2.00-3.00
	3	1.00-2.00	3	1.00-2.00
	4	<1.00	4	<1.00
Witch flounder	1	>0.50	1	>0.50
	2	0.30-0.50	2	0.30-0.50
	3	0.10-0.30	3	0.10-0.30
Wolffish	1	>3.00	1	>3.00
	2	1.00-3.00	2	1.00-3.00
	3	<1.00	3	<1.00
All species	**9**	**Unsorted**	**9**	**Unsorted**

Table 2. Vessel ID, remarks and whether access to SIF data has been granted for contacted vessels. 4.a = Northern North Sea, 4.b = Central North Sea, 3.a = Skagerrak and Kattegat, 22-28 = Baltic Sea. Vessels where owners were unwilling to share SIF or who are undecided have been aggregated into groups based on reason for not granting access or remark on current status.

Vessel	Access granted	Usable haul data	Main fishing areas	First entry at haul level	Remarks
A	Yes	Yes	4.a, 4.b, 3.a	10-04-2015	Number of hauls in SIF 2015-2016: 1473
B	Yes	Yes	4.a, 4.b, 3.a	27-03-2014	Number of hauls in SIF 2015-2016: 925
C	Yes	Yes	4.a, 4.b, 3.a, 22-28	09-12-2013	Number of hauls in SIF 2015-2016: 928
D	Yes	Yes	4.a, 4.b, 3.a	20-03-2015	Number of hauls in SIF 2015-2016: 1418
E	Yes	Yes	4.a, 4.b, 3.a	19-12-2013	Number of hauls in SIF 2015-2016: 1062
F	Yes	Yes	4.a, 4.b, 3.a	19-10-2016	Number of hauls in SIF 2015-2016: 118
N1	No				Believe it to be too expensive in time and money to look into their SIF data
N2, N3	No				No reason given
N4,N5	No				Only sea-pack hake. Did not see the use of sharing the data for one species
N6-N10	No				Use the sea-packing machinery to clean the fish and report to the eLog.
N11	No				Was uncertain as to whether the data could be misused
U1-U4	Undecided				Waiting for email confirmation
Q	Yes	No	4.b, 3.a, 22-28	None	Only sales slips records in SIF.
T	Yes	No	4.a, 4.b, 3.a	08-05-2012	Stop sea-packing in January 2015 due to change in vessel ownership
V	Yes	No	4.b	None	Gillnetter. No hauls. Sea-packing is recorded at day level.
W	Yes	No	4.b, 3.a	05-12-2013	Use the sea-packing machinery to clean the fish and report to the eLog.
X	Yes	No	4.a, 4.b, 3.a, 22-28	20-12-2013	Manually enter haul positions and time. Haul positions and timestamps are unreliable
Y	Yes	No	4.a, 4.b, 3.a, 22-28	17-12-2013	Use the sea-packing machinery to clean the fish and report to the eLog.
Z	Yes	No	4.a, 4.b, 3.a	02-12-2013	Use the sea-packing machinery to clean the fish and report to the eLog.

Table 3. Completeness of SIF when compared to the eLog (hauls and trips) and vessel landings data from DFAD for the 10 most landed species in 2015 and 2016.

	Completeness [%]									
	Vessel A		**Vessel B**		**Vessel C**		**Vessel D**		**Vessel E**	
Species	***2015***	***2016***	***2015***	***2016***	***2015***	***2016***	***2015***	***2016***	***2015***	***2016***
Wolffish	81.1	94.9	87.7	83.6	49.6	60.5	66.4	75.0	62.5	85.5
Lemon sole	88.0	77.9	77.2	100.0	58.7	67.9	41.4	54.6	63.0	86.3
Witch flounder	91.8	89.7	96.0	91.8	46.6	51.8	59.0	52.9	61.2	81.6
Hake	95.2	87.1	90.0	93.0	57.5	64.4	51.1	69.9	69.0	77.1
Turbot	79.0	82.4	93.3	76.3	58.6	68.8	16.1	76.8	64.7	83.2
Haddock	81.4	88.9	96.8	85.3	52.0	69.2	51.8	69.4	62.6	70.6
Monkfish	94.2	91.1	95.3	90.2	60.5	59.6	56.8	73.2	58.9	76.3
Cod	85.0	89.3	93.9	89.4	20.2	29.4	62.6	77.4	63.4	77.3
Saithe	68.0	94.7	91.8	90.7	21.5	55.7	60.7	70.3	55.3	74.6
Plaice	19.1	15.5	90.0	96.4	56.3	64.2	45.6	63.8	61.6	84.3
Overall species results	**78.3**	**81.2**	**91.2**	**89.0**	**48.2**	**59.2**	**51.2**	**68.3**	**62.2**	**79.7**
Fishing trips, number in SIF	**39**	**67**	**35**	**37**	**83**	**88**	**59**	**53**	**42**	**48**
Fishing trips, completeness [%]	**100.0**	**100.0**	**89.7**	**78.7**	**98.8**	**100.0**	**100.0**	**98.1**	**95.5**	**100.0**
Hauls, completeness [%]	**89.8**	**82.6**	**92.3**	**74.8**	**82.6**	**71.5**	**61.6**	**79.0**	**65.3**	**80.0**

Table 4. Mean and standard deviation in percentage of size classes as well as p-value from Wilcoxon rank-sum test. Comparison is done solely using DFAD data between trips where only DFAD data exist and trips where both SIF and DFAD data exist. *Vessel A is not included for monkfish as the vessel do not sea-pack monkfish.

Species	Size class	p-value	Mean percent ± SD [%] only DFAD	Mean percent ± SD [%] SIF and DFAD
Cod	1	0.875	22.7 ± 28.8	15.3 ± 14.1
	2	0.276	26.3 ± 17.6	25.6 ± 12.3
	3	0.002	35.8 ± 17.0	29.9 ± 15.1
	4	0.006	28.1 ± 19.5	22.1 ± 11.5
	5	0.167	20.1 ± 27.6	12.6 ± 12.8
Hake	1	0.004	36.6 ± 25.9	38.2 ± 17.8
	2	0.999	54.4 ± 25.6	54.2 ± 20.1
	3	0.004	34.2 ± 34.2	14.4 ± 14.1
	4	0.080	33.4 ± 34.5	10.1 ± 11.7
Plaice	1	0.007	32.1 ± 29.1	24.0 ± 25.4
	2	0.520	29.1 ± 16.2	28.6 ± 12.6
	3	0.178	29.1 ± 18.0	30.4 ± 13.2
	4	0.821	31.6 ± 26.7	27.8 ± 17.5
Haddock	1	0.006	43.8 ± 24.2	35.7 ± 20.9
	2	0.006	52.9 ± 24.5	51.4 ± 18.4
	3	0.082	30.7 ± 27.4	26.3 ± 15.7
	4	0.707	34.6 ± 40.0	9.0 ± 6.3
Saithe	1	0.056	30.9 ± 31.4	23.7 ± 28.6
	2	< 0.001	35.6 ± 33.9	16.6 ± 17.1
	3	0.234	40.6 ± 27.7	42.0 ± 22.7
	4	0.049	55.8 ± 27.5	46.8 ± 27.3
Lemon sole	1	0.072	24.1 ± 23.8	16.3 ± 12.6
	2	0.595	60.7 ± 16.4	60.3 ± 14.2
	3	0.038	34.5 ± 22.4	28.4 ± 13.9
*Monkfish**	1	0.138	23.1 ± 25.8	15.0 ± 10.9
	2	0.186	25.0 ± 14.4	21.9 ± 10.0
	3	0.807	37.2 ± 15.7	36.9 ± 11.6
	4	0.004	34.9 ± 22.6	27.0 ± 14.1
	5	< 0.001	27.1 ± 33.7	10.1 ± 14.5
Turbot	1	0.820	36.2 ± 32.6	35.1 ± 30.9
	2	0.083	34.6 ± 27.8	27.5 ± 20.4
	3	0.401	51.9 ± 24.9	48.5 ± 22.3
	4	0.013	30.8 ± 26.5	20.9 ± 15.6
Witch flounder	1	0.889	30.8 ± 24.5	28.2 ± 18.3
	2	0.012	67.4 ± 24.4	59.9 ± 19.6
	3	0.331	35.9 ± 27.3	28.7 ± 15.7
Wolffish	1	0.940	52.7 ± 24.1	52.2 ± 22.6
	2	< 0.001	70.0 ± 28.7	58.4 ± 26.6
	3	< 0.001	50.6 ± 44.9	6.3 ± 6.3

Table 5. R^2 and degrees of freedom for linear model fit of landings in SIF and DFAD for the 10 most landed species in 2015 and 2016. SIF data has been aggregated to trip level in order to make the comparison possible with DFAD and comparison is done solely for trips where both SIF and DFAD have records.

	Vessel A				**Vessel B**				**Vessel C**				**Vessel D**				**Vessel E**			
	2015		*2016*		*2015*		*2016*		*2015*		*2016*		*2015*		*2016*		*2015*		*2016*	
Species	**df**	**R^2**	**df**	**R^2**	**df**	**R^2**	**df**	**R^2**	**df**	**R^2**	**df**	**R^2**	**df**	**R^2**	**df**	**R^2**	**df**	**R^2**	**df**	**R^2**
Wolffish	38	0.997	82	0.993	35	0.999	59	0.885	40	0.793	65	0.952	46	0.946	60	0.953	62	0.914	81	0.975
Lemon sole	88	0.936	156	0.978	5	0.574	8	0.859	146	0.836	155	0.985	56	0.981	100	0.890	65	0.944	88	0.985
Witch flounder	55	0.966	84	0.975	33	0.986	12	0.805	28	0.952	69	0.999	21	0.995	83	0.876	82	0.841	105	0.912
Hake	38	0.979	39	0.985	37	0.987	65	0.997	77	0.747	77	0.981	53	0.701	71	0.777	94	0.775	131	0.963
Turbot	117	0.946	228	0.921	27	0.899	31	0.732	120	0.940	165	0.949	30	0.988	79	0.940	40	0.919	63	0.983
Haddock	40	0.972	89	0.855	50	0.991	26	0.704	95	0.880	111	0.978	72	0.813	95	0.552	98	0.831	139	0.857
Monkfish	NA	NA	NA	NA	69	0.997	121	0.996	139	0.933	191	0.886	75	0.922	161	0.899	152	0.749	181	0.880
Cod	122	0.776	212	0.981	56	0.994	109	0.998	227	0.703	260	0.713	125	0.702	169	0.743	160	0.607	182	0.803
Saithe	22	0.908	27	0.994	56	0.999	98	0.998	13	0.737	63	0.963	61	0.904	40	0.853	123	0.731	146	0.918
Plaice	70	0.524	124	0.472	9	0.825	5	0.779	201	0.673	207	0.763	49	0.889	104	0.889	81	0.897	94	0.980

Table 6. ANCOVA output for the effect of year, vessel and size class as well as remaining effect of log-transformed landings from DFAD and residuals.

Species	**Term**	**df**	**Sum of Squares**	**F-value**	**p-value**
Cod	log(DFAD)	1	65411.5	$1.701*10^5$	< 0.001
	Size class	5	10.6	5.490	< 0.001
	Vessel	4	35.5	23.083	< 0.001
	Year	1	2.1	5.441	0.019
	Residuals	1902	731.5		
Hake	log(DFAD)	1	26669.1	91823.644	< 0.001
	Size class	5	2.3	1.591	< 0.001
	Vessel	4	12.9	11.144	< 0.001
	Year	1	1.8	6.207	< 0.001
	Residuals	777	225.7		
Plaice	log(DFAD)	1	41100.6	86700.459	< 0.001
	Size class	5	18.6	7.831	< 0.001
	Vessel	4	14.9	7.874	< 0.001
	Year	1	1.7	3.581	0.059
	Residuals	1019	483.1		
Haddock	log(DFAD)	1	20513.9	64726.260	< 0.001
	Size class	5	1.5	4.579	< 0.001
	Vessel	4	0.4	1.398	0.233
	Year	1	0.1	0.258	0.611
	Residuals	863	273.5		
Saithe	log(DFAD)	1	28087.3	117068.100	< 0.001
	Size class	5	92.5	0.771	< 0.001
	Vessel	4	15.4	16.000	< 0.001
	Year	1	4.8	20.000	0.571
	Residuals	736	176.6		
Lemon sole	log(DFAD)	1	20182.6	183262.301	< 0.001

	Size class	4	3.7	8.292	< 0.001
	Vessel	4	2.1	4.694	< 0.001
	Year	1	1.5	13.782	< 0.001
	Residuals	905	99.7		
Monkfish	log(DFAD)	1	29525.7	145400.713	< 0.001
	Size class	6	64.4	52.859	< 0.001
	Vessel	4	70.3	86.536	< 0.001
	Year	1	0.1	0.356	0.551
	Residuals	1208	245.3		
Turbot	log(DFAD)	1	10539.5	108462.110	< 0.001
	Size class	5	3.8	7.881	< 0.001
	Vessel	4	0.8	2.145	0.073
	Year	1	0.1	0.082	0.774
	Residuals	930	90.4		
Witch flounder	log(DFAD)	1	12335.2	110089.201	< 0.001
	Size class	4	1.0	2.213	0.066
	Vessel	4	2.2	4.965	< 0.001
	Year	1	0.1	0.718	0.397
	Residuals	626	70.1		
Wolffish	log(DFAD)	1	10820.2	93123.94261	< 0.001
	Size class	4	2.3	4.852	< 0.001
	Vessel	4	1.9	4.182	0.002
	Year	1	0.2	1.464	0.228
	Residuals	595	69.1		

Table 7. Mean latitude and longitude as well as p-value from Wilcoxon rank-sum test for all hauls recorded in SIF and EM during 2015 and 2016. Two vessels had records in both datasets. Due to confidentiality agreements, the number of hauls cannot be revealed, however it exceeded 1000 observations in both years.

Year	Mean latitude, SIF	Mean latitude, EM	p-value	Mean longitude, SIF	Mean longitude, EM	p-value
2015	58.16°N	58.26°N	<0.001	4.72°E	4.34°E	<0.001
2016	58.17°N	58.27°N	0.174	4.71°E	4.34°E	0.701

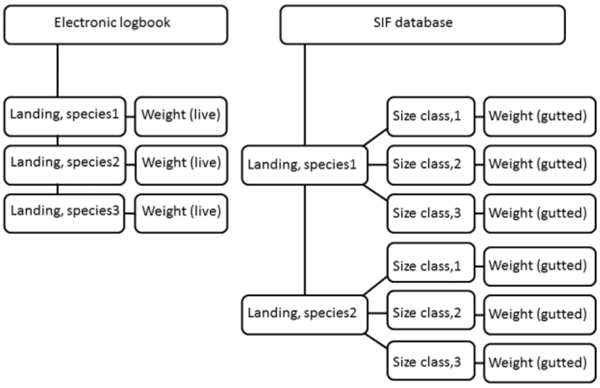

Figure 1. Conceptual figure of the difference between landings data available at haul level in the electronic logbook and the sea-packing data available in the SIF database.

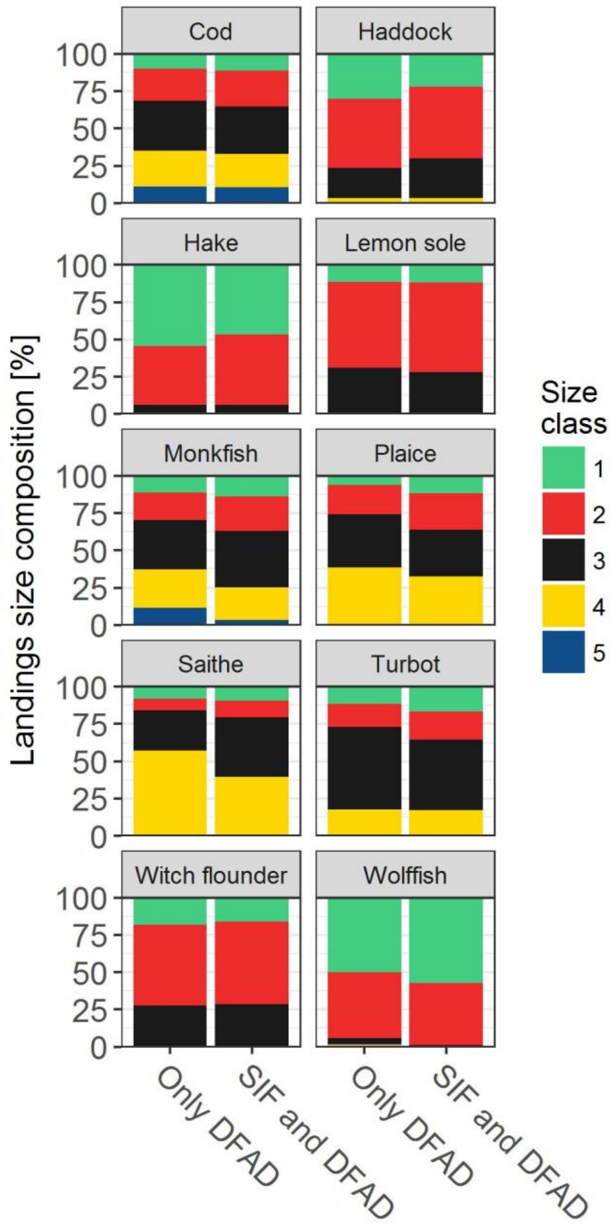

Figure 2. Landings' size composition in percent stratified on trips with only DFAD data and trips with both DFAD and SIF data. Size class 1 are the largest specimens.

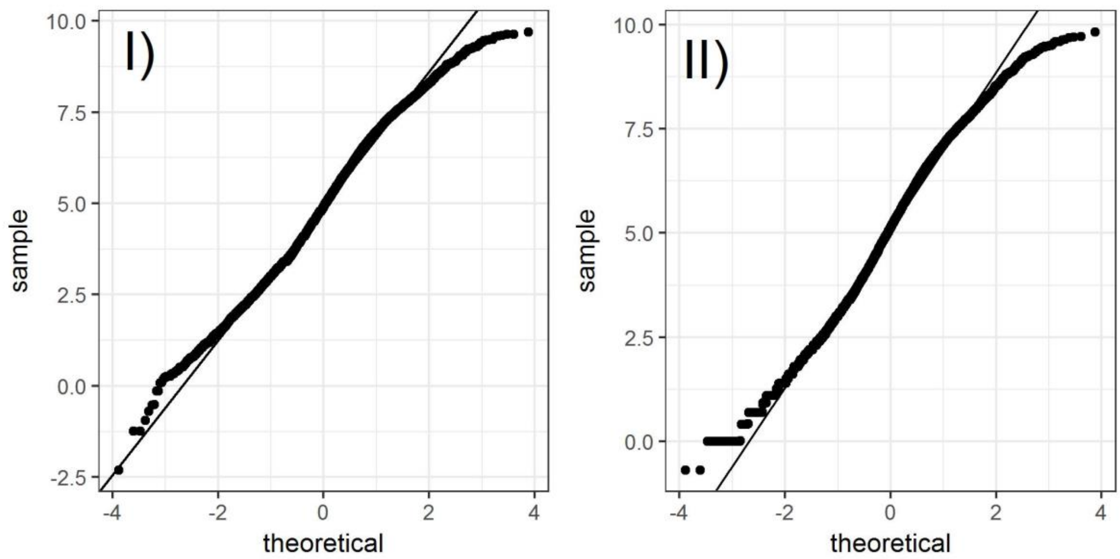

Figure 3. QQ-plot for I) log-transformed landings recorded in SIF. II) log-transformed landings recorded in DFAD.

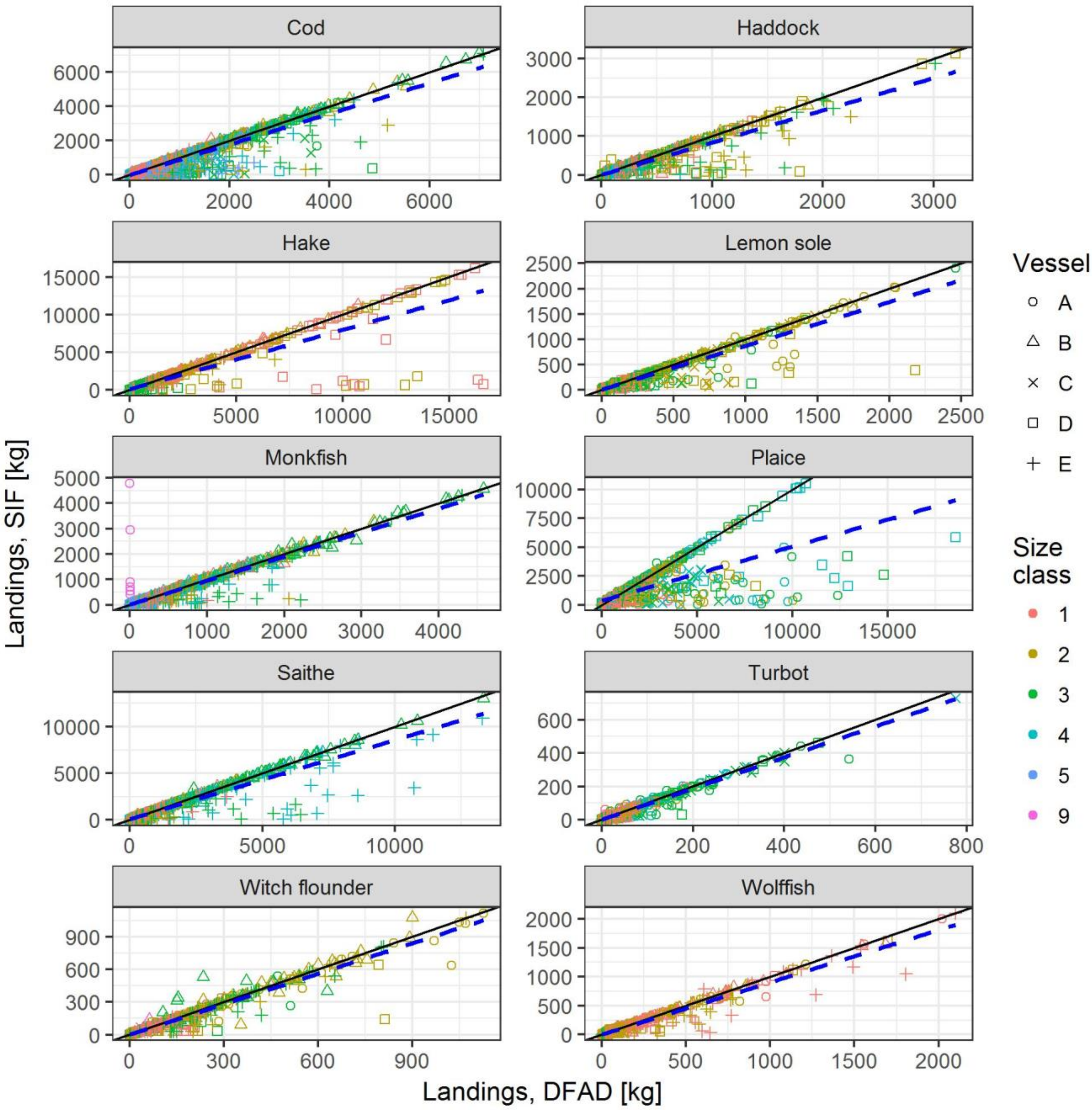

Figure 4. Landings per trip according to DFAD and SIF for the 10 most landed species in 2015 and 2016 by species and commercial size class. Points: The aggregated weight of the species and size class for a fishing trip. The x-axis represent the weight according to DFAD, the y-axis represent the weight according to SIF. Blue dashed line: linear model fit between DFAD and SIF. Black line: The 1:1 ratio between DFAD and SIF. Size class 9 is unsorted.

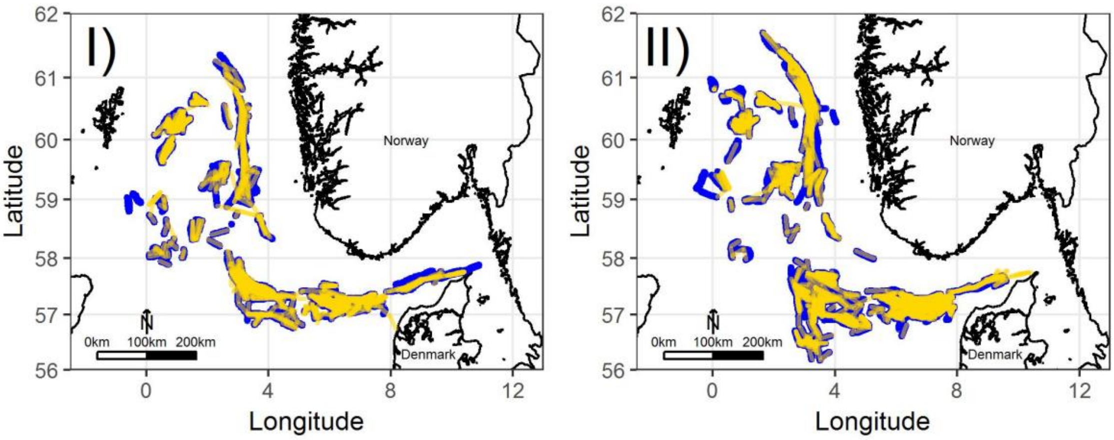

Figure 5. Fishing activity overlap between EM and SIF for two vessels. I) 2015. II) 2016.Blue points: Fishing activity recorded by EM GPS sensors (1-minute interval). Yellow lines: Hauls according to SIF. The EM trial did not cover the Baltic Sea and the maps do therefore not include hauls in this area.