ICES Journal of Marine Science (2011), 68(8), 1712–1718. doi:10.1093/icesjms/fsr066

A method of detecting patterns in mean lengths of samples of discarded fish, applied to the self-sampling programme of the Dutch bottom-trawl fishery

Sebastian S. Uhlmann*, Stijn M. Bierman, and Aloysius T. M. van Helmond

Wageningen Institute of Marine Resources and Ecosystem Studies (IMARES), PO Box 68, 1970 AB IJmuiden, The Netherlands

**Corresponding Author: tel: +31 317 480133; fax: +31 317 487326; e-mail: sebastian.uhlmann@wur.nl.*

Uhlmann, S. S., Bierman, S. M., and van Helmond, A. T. M. 2011. A method of detecting patterns in mean lengths of samples of discarded fish, applied to the self-sampling programme of the Dutch bottom-trawl fishery. – ICES Journal of Marine Science, 68: 1712–1718.

Received 22 October 2010; accepted 28 March 2011; advance access publication 8 June 2011

In 2009, a self-sampling programme was organized in the Netherlands, fishers sampling ca. 80 kg of discards from randomly selected bottom trawls in the North Sea. A statistical procedure is proposed to highlight samples, trips (with multiple samples), or vessels (which may have multiple trips within a year) where extreme mean lengths of discarded fish were observed. Randomization methods were used to test for evidence of non-randomness in patterns of highlighted discard samples, e.g. repeated observations of extreme mean lengths for consecutive discard samples across trips from the same vessel. European plaice (*Pleuronectes platessa*), common dab (*Limanda limanda*), grey gurnard (*Eutrigla gurnardus*), and whiting (*Merlangius merlangus*) were considered because these were the most abundant species in most of the discard samples. A linear mixed model was used to estimate random-sample effects on the estimated mean lengths by species. These random effects were incorporated into uni- and bivariate procedures to identify extreme samples that were summed for each vessel, and the probability of observing such numbers was estimated. Excluding these samples from the dataset had marginal effects on estimated size distributions of fish.

Keywords: at-sea sampling, data quality, discards, self-reporting.

Introduction

At-sea sampling of commercial fish catches by observers is expensive because the observers typically have to remain on board for the duration of a trip. This tends to return large clusters of samples from a few trips, which may lead to small effective sample sizes (e.g. Pennington and Vølstad, 1994), when the aim is to make inferences for all trips made by the whole fleet. From this perspective, self-sampling by fishers is an attractive alternative because more samples from more trips can be collected with unit costs being lower. Compared with the long-term fishery-observer programme organized under the European Data Collection Framework (EU Regulations 1543/2000 and 10121/2009), the benefit has been demonstrated for a self-sampling programme conceived at the Institute for Marine Resources and Ecosystem Studies (IMARES, Wageningen University; see van Helmond and van Overzee, 2010, for detail). In both programmes, apart from general biological, technical, and environmental information, length frequency data are collected for discards of the Dutch bottom-trawl fishery in the North Sea. Ideally, these data are used for stock assessment.

However, fishery-dependent length frequency data may be biased by systematic sampling errors that can influence stock assessments seriously (Heery and Berkson, 2009). Self-sampling may be particularly prone to such bias, because fishers routinely and subjectively select fish from the catch during their daily commercial operations (sorting ogive), but potentially non-randomly subsample the discards for subsequent biological analysis (sampling ogive). Fishers may find it difficult to conform to the more objective sampling regime required for scientific monitoring. Although sorting ogives may be similar across vessels, especially when targeting species with a minimum landing size (MLS; Appendix XII of EC Council Regulation No 850/98), sampling ogives may differ, especially if fishers consistently and non-randomly pick and/or miss certain size classes of a species.

Lacking any independent *in situ* validation techniques (e.g. video-assisted monitoring; Ames *et al.*, 2007; Stanley *et al.*, 2009), a *post hoc* statistical screening method is developed here to detect patterns in the mean lengths of samples of discarded fish across species, hauls, vessels, and trips which may suggest biased sampling at a haul level. Self-reported data may also be biased at the sorting level as a consequence of fishers misreporting catches and/or discards to circumvent regulations, e.g. on quota and MLS (Bremner *et al.*, 2009; Heery and Berkson, 2009; Bousquet *et al.*, 2010). This can arise with large marketable fish or small fish (below MLS); in either case, the sampled size distribution of the discards will be biased.

Historically, this problem has been observed in comparisons of the discard fractions of European plaice (*Pleuronectes platessa*) and Atlantic cod (*Gadus morhua*) reported from observer and self-sampling operations in the Dutch beam-trawl fishery (Aarts and van Helmond, 2007). The different length frequency distributions for plaice, despite accounting for spatial and temporal effects,

© 2011 International Council for the Exploration of the Sea. Published by Oxford Journals. All rights reserved.
For Permissions, please email: journals.permissions@oup.com

Downloaded from https://academic.oup.com/icesjms/article-abstract/68/8/1712/750498 by guest on 18 February 2019

suggested that discarded small fish were consistently missing from the samples (this term is used here instead of "underreporting", because the latter implies a deliberate process, which it may not be) in the self-sampling programme (Aarts and van Helmond, 2007). Because of these discrepancies, the data from this self-sampling programme were considered unsuitable for stock assessments.

Since the study of Aarts and van Helmond (2007), the self-sampling programme has shifted from an industry-driven initiative (designed and organized by staff of the Dutch Fish Product Board, from 2004 to 2008) to a scientific sampling scheme (designed, organized, and analysed by IMARES staff, from 2009 on) which has operated in parallel with the long-term observer programme. In the current IMARES self-sampling programme, there is a reference fleet ($n = 12$ vessels in 2009) with trained observers among the crew who opportunistically and voluntarily collect discard samples during commercial fishing operations throughout the year. In accord with the instructions of IMARES staff, two random and pre-determined hauls are sampled on an agreed trip. One sample comprises two boxes of discards (a box weighs ca. 40 kg), filled by taking subsamples ideally at intervals while the catch is sorted (Heales *et al*., 2003). For each sampled haul, additional information on the composition and volume of catch and landings, environmental factors (e.g. wind direction and speed, latitude and longitude, and water depth) and operational details (e.g. start and end times of trawling, gear type, and mesh size) are also recorded. All discard samples are returned to the laboratory where the species composition, size, and age structure of the sample is determined. European plaice, common dab (*Limanda limanda*), grey gurnard (*Eutrigla gurnardus*), and whiting (*Merlangius merlangus*) are among the most commonly discarded species (van Helmond and van Overzee, 2010).

Here, we present a statistical tool to highlight samples, trips (with multiple samples), or vessels (with multiple trips) for which (i) the on-board sorting into discards and landings, (ii) the on-board sampling of individual fish from the discard fraction for return to the laboratory, or both have led to mean length in a sample being different from other samples. Process (ii) may indicate sampling bias. However, our statistical tool cannot establish which of processes (i) or (ii) prevails, especially for species without an MLS. It can, however, visualize simultaneous occurrences of extreme values. Notwithstanding this, the tool can be used for rapid assessment of potential biases in the estimated mean fish lengths of discards by species where each sample is taken at a haul level. Because of the geographic spread of sampling, different populations of discarded fish are sampled by the observer and self-sampling programmes (Figure 1). Therefore, the present study focuses on the data from the Dutch self-sampling programme in 2009, as a case study.

Material and methods

The numbers-at-length of discarded European plaice, common dab, grey gurnard, and whiting were extracted from the IMARES database. Samples, i.e. two boxes (ca. 80 kg) of discards per haul, were returned from two fleet segments each with two characteristic mesh sizes (in total, four métiers) operating in ICES Divisions IVc and IVb throughout the year, namely beam and otter trawlers with 80 and 100 mm mesh sizes. Discards were sampled from 133 hauls on a total of 70 trips in each month of 2009. For each haul, the numbers-at-length were raised to the haul level, based on the fraction of the subsample, i.e. two boxes out of the total number of boxes discarded. All data were checked carefully for transcription errors and missing values.

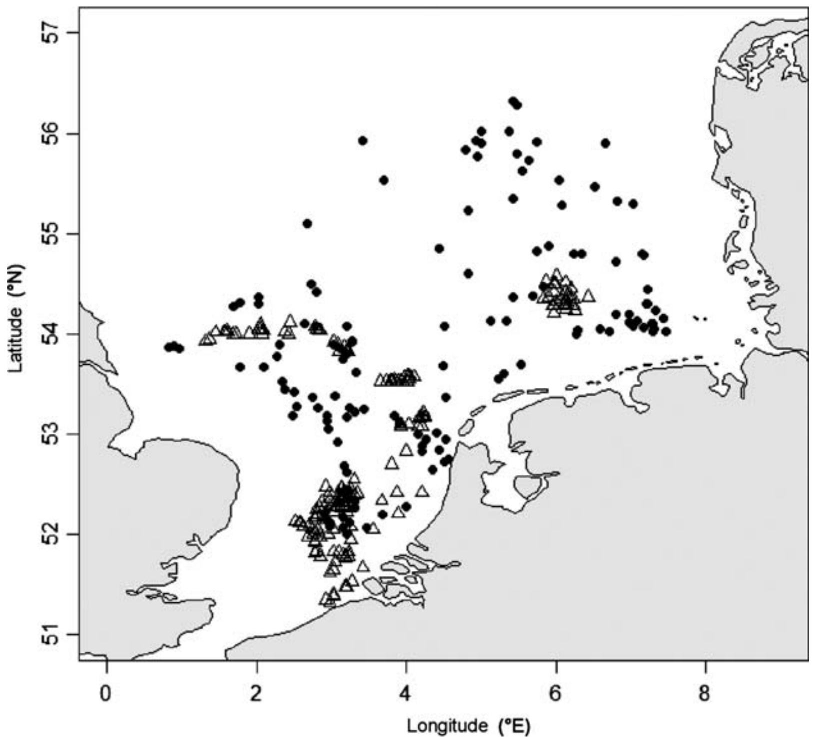

Figure 1. Geographic locations of hauls sampled in 2009 for the Dutch bottom-trawl fishery by the observer (open triangles) and self-sampling (dots) programmes.

Downloaded from https://academic.oup.com/icesjms/article-abstract/68/8/1712/750498 by guest on 18 February 2019

Statistical analysis

Mixed model for estimating random-sample effects on mean lengths

The means of the measured discarded fish lengths by species were expected to vary as a result of changes in the underlying population from which the catch was taken, the selectivity of the gear, the on-board sampling method, and sorting and sampling ogives (Benoît and Allard, 2009). Therefore, we modelled the expected mean fish lengths in the absence of on-board sorting and sampling bias as a function of location, season, and gear type. Location was treated as three distinct areas to reflect the distribution of the metiérs, e.g. mesh sizes >100 mm need to be used north of 55°N: ≥51 to <53.5°N; ≥53.5° to 55°N; and ≥55°N. The number of measured fish per species in a sample (corresponding to a haul) can vary from just 1 to >100. We chose a mixed-model approach in which sample effects on mean lengths are estimated as random effects, because in that case the estimated sample effects based on a few fish will decrease towards the expected mean length (Gelman *et al.*, 1995).

Let y_{ji} be the measured length of fish i ($i = 1,2,\ldots, n_j$) in sample j, where n_j is the number of measured fish in sample j. For readability, we do not use a subscript for species here; the same model applies to each species. Then, a random-sample effect can be estimated using the following mixed model:

$$\begin{aligned} y_{j,i} = {} & \alpha + \beta_1 \text{gear}_{g(j)} + \beta_2 \text{area}_{a(j)} + \beta_3 \text{quarter}_{q(j)} \\ & + \beta_4 \text{area}_{a(j)} \times \text{quarter}_{q(j)} + r_{r(j)} + g_j + \varepsilon_{ij}, \end{aligned} \tag{1}$$

where $\text{gear}_{g(j)}$, $\text{area}_{a(j)}$, and $\text{quarter}_{q(j)}$ are fixed-effect parameters for gear type g, area a ($a \times \{1,2,3\}$), and quarter q ($q \times \{1,2,3,4\}$), corresponding to sample j, and $\text{area}_{a(j)} \times \text{quarter}_{q(j)}$ is the interaction between these factors. Random effects are $r_{r(j)}$ for the combination of quarter and ICES rectangle r in which sample j was taken, i.e. accounting for the between-rectangle variability within a given area, and g_j are random-sample effects. The residual error term is ε_{ij} for fish i in sample j. Both random effects are assumed to be normally distributed with a mean of zero and variances σ_r^2 and σ_g^2, respectively. The distribution of length measurements was also modelled by a normal distribution (error term).

Uni- and bivariate approaches

Extreme values (with reference to a normal distribution) of random-sample effects g_j as estimated using the mixed model, Equation (1), may indicate a different sorting ogive or a sampling bias, particularly if large/small values of g_j were estimated simultaneously for multiple species within the same trip (across hauls) or for multiple trips by the same vessel. To investigate this, we counted the number of extreme values in the estimated random-sample effects per trip and vessel, taking both univariate (per species) and bivariate (with combinations of species) approaches. Although the latter approach could extend to many more dimensions, two seemed appropriate here, because including more species would result in too few samples per category to be useful. We chose to couple the two most abundant species groups (European plaice and common dab; and grey gurnard and whiting, respectively) because most samples had at least one measured fish of each of these species. Univariately, results were classified beyond the 2.5 and 97.5 percentiles of the random-sample effects by species as extreme. The choice of percentile is subjective and arbitrary and can be varied by the analyst. For the univariate and bivariate methods, percentiles need to be selected to return numbers of extreme samples that are neither too small nor too large to identify patterns and to compute p-values using the randomization method.

Bivariately, the distance–distance plot methodology proposed by Rousseeuw and van Zomeren (1990) was used to classify extreme samples in bivariate space, based on comparing a robust version of the Mahalanobis distance with the quantiles of the Chi-squared distribution, with 2 degrees of freedom (Garrett, 1989). This classification method circumvents potential problems with biased estimation of the multivariate mean and covariance matrix attributable to the presence of potential extreme values, based on the minimum covariance determinant (MCD) estimator of Rousseeuw and van Driessen (1999). As the random-sample effects are estimated independently for each species, the multivariate mean may not necessarily be at zero.

Finally, using the bivariate extreme samples from the European plaice and common dab group, a randomization test (Manly, 2007) was used to investigate whether the observed numbers of extreme samples per vessel could have occurred by chance. In all, 5000 replicate datasets were simulated by randomly reordering the flags (extreme sample or not) across all samples. For each replicate dataset, the number of flags per vessel was counted, and the number of flags per vessel compared with the observed number of flags per vessel, to estimate the chance of observing the same number or more flags. Bonferroni correction (Gotelli and Ellison, 2004) was applied to account for the multiplicity of tests if more than one vessel had flagged samples.

To illustrate how the estimated length distribution of discarded European plaice and common dab changed by excluding the extreme samples identified in the bivariate approach, relative length frequency distributions (i.e. proportions per size class) for these species were plotted from all self-sampled trips in 2009. The size frequency distributions (at 1-cm intervals) of counts (raised to trip level) of European plaice and common dab, from samples including or excluding extreme samples, were compared using two-sample Kolmogorov–Smirnov tests.

The mixed-model analyses were carried out using the statistical software R (R Development Core Team, 2005), with the aid of the "ellipse" (Murdoch and Chow, 1996) and "mvoutlier" (Filzmoser *et al.*, 2005) packages, which contain routines for drawing ellipse-like confidence regions, and estimation of robust Mahalanobis distances using the MCD method for estimating variance–covariance matrices. The package "nlme" (Pinheiro *et al.*, 2009) was used to fit the random-effects model. All packages can be downloaded from http://cran.r-project.org.

Results

For the univariate method of classifying extreme samples (using the random-sample effects on a per-species basis), 130 samples with measured fish were included (European plaice, $n = 127$; common dab, $n = 130$; grey gurnard, $n = 109$; whiting, $n = 89$; Table 1, Figure 2).

All but one of the 12 vessels participating in the self-sampling programme in 2009 returned at least one sample with either a positive or a negative sample effect (estimated mean lengths greater or smaller than expected) for at least one of the species measured (Table 1). Within any sample, no more than two extreme mean lengths across the four species were evident (Table 1); more extreme mean lengths were found for European plaice and common dab (Table 1). Within a trip up to three,

Downloaded from https://academic.oup.com/icesjms/article-abstract/68/8/1712/750498 by guest on 18 February 2019

Table 1. List of vessel codes, number of sampled trips (n), and sample codes for which at least one random-sample effect for plaice, dab, grey gurnard, and whiting was classified as extreme (univariate method; see Figure 2).

Vessel code	n	Trip code	Sample code	Plaice	Dab	Grey gurnard	Whiting
1	4	119	6000684	0	+	+	n/a
2	9	124	6000602	−	0	0	0
2	9	126	6000629	−	0	0	0
2	9	127	6000679	−	0	0	n/a
2	9	128	6000700	+	0	0	0
2	9	130	6000725	0	0	0	+
2	9	130	6000726	0	0	0	−
3	2	134	6000685	0	+	+	n/a
4	2	135	6000609	0	0	n/a	−
4	2	136	6000643	+	0	n/a	0
5	8	138	6000623	0	−	0	0
5	8	138	6000624	0	0	0	+
5	8	140	6000663	0	0	0	+
7	8	149	6000662	0	−	0	n/a
8	9	155	6000605	0	0	+	0
8	9	156	6000632	0	+	0	n/a
8	9	157	6000659	+	0	0	n/a
9	8	167	6000670	0	0	−	n/a
10	8	173	6000612	−	0	0	0
11	5	182	6000647	+	+	0	n/a
12	6	187	6000636	0	−	0	−
12	6	189	6000707	0	0	−	0
12	6	189	6000708	0	−	−	0
$n = 11$	69	20	23	4+/4−	4+/4−	3+/3−	3+/3−

The extreme cases are shown as − or + for, respectively, extreme negative or positive random-sample effects, and 0 for all others. The total number of samples for each category "(n; vessel, trip, sample, and positive/negative random-sample effect per species) are given in the bottom row. n/a, no data available.

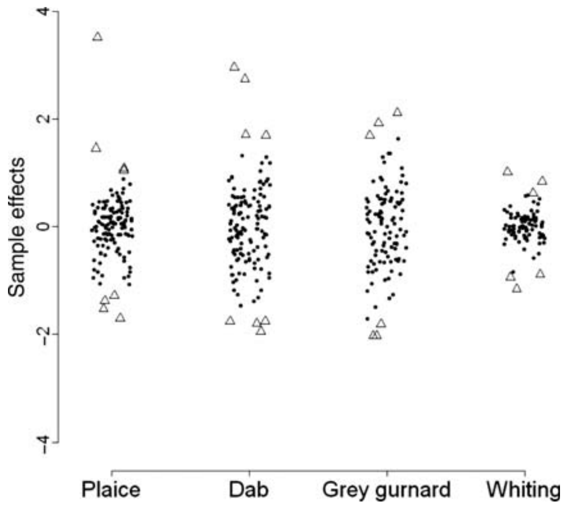

Figure 2. Classification of extreme length measurements using the univariate approach. The smallest (<2.5 percentile) and the largest (>97.5 percentile) of the random-sample effects estimated using the mixed model [Equation (1)] for plaice, dab, grey gurnard, and whiting are deemed extreme (triangles); other data points are shown as dots.

and within a vessel up to six, extreme mean lengths were recorded. Of these, four extreme negative mean lengths were returned for a particular vessel (code "2"; three and one for MLS-regulated European plaice and whiting, respectively), although overall the numbers of positive or negative sample effects were evenly distributed within and across species (Table 1).

For the bivariate method (excluding extreme values), 126 samples with measurements of both European plaice and common dab could be included, along with 69 with both grey gurnard and whiting (Table 2, Figure 3).

Sample effects (extreme values) were flagged for data collected on 8 of the 12 vessels (Table 2). For the European plaice and common dab group, five and three samples were flagged as falling outside the 95 and 99% prediction intervals of the normal random effect, respectively (Figure 3a). For grey gurnard and whiting, the corresponding numbers were five and one samples, respectively (Figure 3c). Notably, for one vessel (code "2" in Tables 1 and 2), samples of both European plaice and common dab were flagged on nearly every trip, and repeatedly in consecutive samples from the same trip (Table 2). This is the same vessel for which the most sample effects were recorded as extreme in the univariate analysis. The number of trips sampled was similar compared with other participating vessels (Table 1). However, given Bonferroni correction ($n = 12$ tests; error rate $p < 0.005$), it appears likely that such a large number of extreme samples could have arisen at least once by chance for a particular vessel if the extreme samples were distributed randomly across all samples (randomization test; Table 3).

There were no significant differences in length frequency distributions of European plaice or common dab whether or not extreme samples identified by the bivariate method were included (Kolmogorov–Smirnov test, $p > 0.05$; Figure 4).

Discussion

Self-sampling programmes are popular (Catchpole and Gray, 2010) because more samples from more trips can be collected at lower cost than during on-board observer programmes. The results here suggest that the length frequencies of self-sampled discards of European plaice, common dab, grey gurnard, and whiting in 2009 provided no evidence that the sampling may have been biased at a vessel level, assuming that all vessels applied the same sorting ogive for discards, because MLS-regulated species were targeted. However, using our uni- and bivariate approaches, we identified individual discard samples (e.g. samples of European plaice from vessel "2"or the top triangles in Figure 3a) that may be considered in greater detail, e.g. by plotting length frequency distributions. Further, we examined the sensitivity of the estimated length-class proportions with and without the trips that returned large random-sample effects, using the bivariate method for European plaice and common dab (Figure 4). Although the variation is negligible, our results may nevertheless be used to identify the crews that need additional training or experience in the sampling methodology or for which it is necessary to study the discard sorting ogive.

Central to the method here is the use of a mixed model to determine random-sample effects on the estimated mean length of discarded fish. An important advantage of the method is that the effects of samples with few measured fish will decrease towards the overall mean of the fixed effects. This avoids the problem that samples with just a few fish might be flagged as extreme. On the other hand, samples with many measured fish may be classified as extreme because the shrinkage effect of the model is less effective in that case. Most samples with a random-sample effect contained at least ten measured fish (Table 2).

Downloaded from https://academic.oup.com/icesjms/article-abstract/68/8/1712/750498 by guest on 18 February 2019

Table 2. List of vessel, trip, and sample codes for which at least one random-sample effect for European plaice, common dab, grey gurnard, and whiting was classified as extreme using the bivariate method, showing the numbers of discarded fish measured, with bivariate 1 (BIV1) and 2 (BIV2) flagging extreme values for discard samples with plaice and dab, and grey gurnard and whiting, respectively (Figure 3).

Vessel code	Trip code	Sample code	BIV1	BIV2	Plaice	Dab	Grey gurnard	Whiting
2	124	6000602	1	0	97	85	4	27
2	126	6000629	1	0	167	106	1	3
2	127	6000679	1	n/a	66	92	1	n/a
2	127	6000680	1	n/a	86	67	n/a	n/a
2	128	6000700	1	0	25	41	5	1
2	130	6000725	0	1	44	14	5	12
2	130	6000726	0	1	104	27	6	7
3	134	6000685	1	n/a	57	13	53	n/a
5	138	6000624	0	1	106	123	3	21
8	160	6000711	1	n/a	50	63	2	n/a
9	170	6000717	1	0	26	56	16	2
10	173	6000612	1	0	17	31	3	5
11	182	6000647	1	n/a	54	50	30	n/a
12	186	6000607	0	1	25	56	39	75
12	187	6000636	1	1	167	77	14	83
$n = 8$	13	15	11	5	1 091	901	182	236

For BIV1 and BIV2,, the extreme values are shown as "1", and "0" otherwise. The total number of samples for each category (*n*; vessel, trip, sample, and random-sample effect per species group) and total number of individual fish measured are given in the bottom row. n/a, no data available.

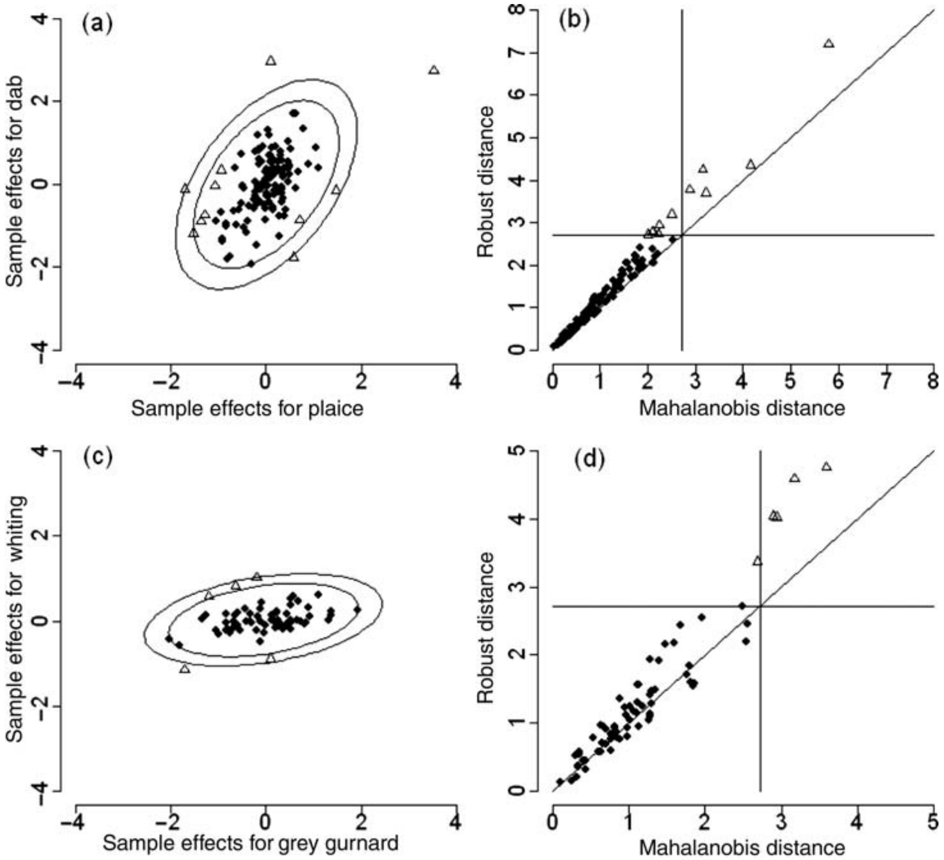

Figure 3. Classification of extreme samples using the bivariate distributions of the random-sample effects estimated using the mixed model [Equation (1)]. The bivariate distribution, with 95 and 99% prediction intervals (inner and outer ellipses, respectively), is shown for (a) plaice vs. dab and (c) grey gurnard vs. whiting. The classification of extreme samples is made using the method of Rousseeuw and van Zomeren (1990) by comparing a robust version of the Mahalanobis distance with the quantiles of the Chi-squared distribution with 2 degrees of freedom. The horizontal and vertical lines in (b) and (d) are drawn at the square roots of the 97.5% quantiles of a Chi-squared distribution with 2 degrees of freedom for (b) plaice and dab, and (d) grey gurnard and whiting. Points above the horizontal line (shown as triangles) are considered extremes.

Downloaded from https://academic.oup.com/icesjms/article-abstract/68/8/1712/750498 by guest on 18 February 2019

Table 3. Results of the randomization test for the number of extreme samples per vessel classified using the bivariate distribution of the random-sample effects for plaice and dab (Figure 3a and b).

	k									
Vessel code	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	***K***	***p*(*k* ≥ *K*)**
1	0.526	0.358	0.094	0.02	0.002	0	0	0	0	1
2	0.203	0.318	0.293	0.134	0.039	0.013	0	0	5	0.013
3	0.668	0.292	0.039	0.001	0	0	0	0	1	0.232
4	0.627	0.312	0.060	0.001	0	0	0	0	0	1
5	0.33	0.393	0.218	0.050	0.007	0.002	0	0	1	0.670
6	0.839	0.155	0.006	0	0	0	0	0	0	1
7	0.303	0.376	0.230	0.079	0.010	0.002	0	0	0	1
8	0.158	0.343	0.318	0.128	0.044	0.008	0	0.001	1	0.842
9	0.249	0.39	0.254	0.079	0.023	0.003	0.002	0	1	0.751
10	0.263	0.403	0.235	0.076	0.020	0.003	0	0	1	0.737
11	0.460	0.399	0.120	0.020	0.001	0	0	0	1	0.54
12	0.334	0.398	0.208	0.051	0.007	0.002	0	0	2	0.286

The probabilities of observing *k* extreme samples per vessel were estimated from 5000 replicate datasets, where the extremes were randomly reordered across samples. The probabilities of observing at least *K* extreme samples per vessel [$p(k \geq K)$] are in the column to the right. The error rate ($p = 0.05$) was divided by the number of hypothesis tests carried out within the randomization analysis (Bonferroni correction, $p < 0.005$).

There are several limitations of the present methodology. First, compared with the univariate method, the bivariate method currently does not identify the direction of the random-sample effect, i.e. positive or negative. Second, classification of individual random-sample effects into extreme or non-extreme values is necessarily partly subjective, influenced to a large extent by the choice of confidence levels. For example, classification based on the 99% prediction intervals (outer ellipses in Figure 3) resulted in fewest samples classified as extreme, whereas the univariate method based on 2.5 and 97.5 percentiles resulted in most (Table 1). Although the choice of confidence level can be varied, the idea behind the methodology is that patterns in highlighted samples are investigated using randomization methods to test for evidence of possible non-randomness in these patterns. Third, the classification of extreme samples relies heavily upon modelling assumptions, so care should be taken in interpreting random-sample effects. Notably, the validity of the method depends upon having a good model for the dependence of sampled mean lengths on the structure of the fish population and the gear-selectivity characteristics. In the mixed model [Equation (1)], these effects were incorporated by including spatial and temporal factors, and their interactions, as well as technical (gear) factors. Another and potentially more robust way of including such effects in the analysis would be to subdivide the data by grouping trips from the same fishing ground, the same season, and the same gear and mesh-size combination. However, in interpreting patterns (if any), one needs to be aware that certain modelling assumptions could have been violated, e.g. that certain explanatory variables were missing or included in the model in the wrong way (e.g. their effect was non-linear when they were included as linear effects). Such misspecifications of the model can introduce bias in the estimated random effects or induce the random effects to be non-normal.

Here, the focus was on detecting potential sampling biases for mean fish length. However, this is just one of several biases that may arise, and alternative important aspects of the sampling and its variance may be looked at using similar methodologies (Vigneau and Mahévas, 2007). The methodology employed is purely statistical and cannot be used to make any inferences on the processes underlying the potential bias in sampling. For that,

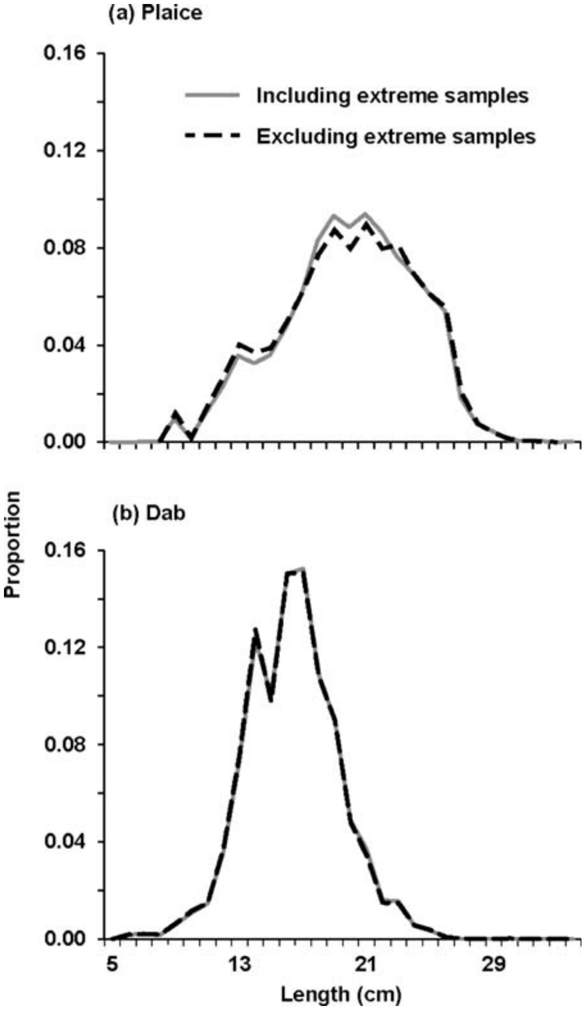

Figure 4. Proportions of the numbers of discarded (a) plaice and (b) dab per trip and size class (cm) from self-sampled discard data for the Dutch bottom-trawl fisheries in 2009. Grey continuous lines, all data included; black dashed lines, length distributions where trips with extreme samples detected by the bivariate method (Table 2 and Figure 3) were excluded.

Downloaded from https://academic.oup.com/icesjms/article-abstract/68/8/1712/750498 by guest on 18 February 2019

less-theoretical, more-practical approaches are needed, such as *in situ* video-monitoring systems to validate logbook catch estimates (Stanley *et al*., 2009), or concurrent sampling by both fishers and on-board observers. Recognizing the importance of having statistical methodology in place to screen data from discard self-sampling programmes, especially considering the incentives for fishers to misreport the occurrence of large marketable and/or small juvenile fish within the discard fraction, so negatively or positively biasing the length frequency distributions, we caution jumping to any foregone conclusion if any extreme samples were to be excluded from a database and/or analysis. Achieving the long-term goal of proving that reliable data can be obtained through self-sampling will eventually promote and maximize the benefits of cooperative research partnerships between fishers, scientists, and managers (Johnson and van Densen, 2007).

Acknowledgements

The work would not have been possible without the dedication of the skippers and crew who participated in the self-sampling programme in 2009. We also thank David MacLennan, Verena Trenkel, and two anonymous referees whose comments greatly improved the manuscript.

References

Aarts, G. M., and van Helmond, A. T. M. 2007. Discard sampling of plaice (*Pleuronectes platessa*) and cod (*Gadus morhua*) in the North Sea by the Dutch demersal fleet from 2004 to 2006. Report C120/07 Prepared for the Dutch Fish Product Board. Institute for Marine Resources and Ecosystem Studies (IMARES), IJmuiden, The Netherlands. 42 pp.

Ames, R. T., Leaman, B. M., and Ames, K. L. 2007. Evaluation of video technology for monitoring of multispecies longline catches. North American Journal of Fisheries Management, 27: 955–964.

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

Bousquet, N., Cadigan, N., Duchesne, T., and Rivest, L. P. 2010. Detecting and correcting underreported catches in fish stock assessment: trial of a new method. Canadian Journal of Fisheries and Aquatic Sciences, 67: 1247–1261.

Bremner, G., Johnstone, P., Bateson, T., and Clarke, P. 2009. Unreported bycatch in the New Zealand West Coast South Island hoki fishery. Marine Policy, 33: 504–512.

Catchpole, T. L., and Gray, T. S. 2010. Reducing discards of fish at sea: a review of European pilot projects. Journal of Environmental Management, 91: 717–723.

Filzmoser, P., Garrett, R. G., and Reimann, C. 2005. Multivariate outlier detection in exploration geochemistry. Computers and Geosciences, 31: 579–587.

Garrett, R. G. 1989. The Chi-square plot: a tool for multivariate outlier recognition. Journal of Geochemical Exploration, 32: 319–341.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. 1995. Bayesian Data Analysis. Chapman and Hall, New York. 526 pp.

Gotelli, N. J., and Ellison, A. M. 2004. A Primer of Ecological Statistics. Sinauer Associates Inc., Sunderland, MA. 510 pp.

Heales, D. S., Brewer, D. T., and Jones, P. N. 2003. Subsampling trawl catches from vessels using seawater hoppers: are catch composition estimates biased? Fisheries Research, 63: 113–120.

Heery, E. C., and Berkson, J. 2009. Systematic errors in length frequency data and their effect on age-structured stock assessment models and management. Transactions of the American Fisheries Society, 138: 218–232.

Johnson, T. R., and van Densen, W. L. T. 2007. Benefits and organization of cooperative research for fisheries management. ICES Journal of Marine Science, 64: 834–840.

Manly, B. J. F. 2007. Randomization, Bootstrap and Monte Carlo Methods in Biology, 3rd edn. Chapman and Hall, Boca Raton, FA. 455 pp.

Murdoch, D. J., and Chow, E. D. 1996. A graphical display of large correlation matrices. The American Statistician, 50: 178–180.

Pennington, M., and Vølstad, J. H. 1994. Assessing the effect of intra-haul correlation and variable density on estimates of population characteristics from marine surveys. Biometrics, 50: 725–732.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and the R Development Core Team. 2009. nlme: linear and nonlinear mixed effects models. R Package, version 3.1-96. http://cran.r-project.org.

R Development Core Team. 2005. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://cran.r-project.org.

Rousseeuw, P. J., and van Driessen, K. 1999. A fast algorithm for the minimum covariance determinant estimator. Technometrics, 41: 212–223.

Rousseeuw, P. J., and van Zomeren, B. C. 1990. Unmasking multivariate outliers and leverage points. Journal of the American Statistical Association, 85: 633–639.

Stanley, R. D., Olsen, N., and Fedoruk, A. 2009. Independent validation of the accuracy of yelloweye rockfish catch estimates from the Canadian groundfish integration pilot project. Marine and Coastal Fisheries: Dynamics, Management, and Ecosystem Science, 1: 354–362.

van Helmond, A. T. M., and van Overzee, H. J. M. 2010. Discard Sampling of the Dutch Beam Trawl Fleet in 2008. Institute for Marine Resources and Ecosystem Studies (IMARES), IJmuiden, The Netherlands. 45 pp. http://www.cvo.wur.nl/default.asp?ZNT=S0T2O-1P316.

Vigneau, J., and Mahevas, S. 2007. Detecting sampling outliers and sampling heterogeneity when catch-at-length is estimated using the ratio estimator. ICES Journal of Marine Science, 64: 1028–1032.

Downloaded from https://academic.oup.com/icesjms/article-abstract/68/8/1712/750498 by guest on 18 February 2019