2016

A field test of fisheries observer sampling methods for estimation of at-sea discards

Jennifer Cahalan
Pacific States Marine Fisheries Commission, Seattle, WA, jcahalan@psmfc.org

Craig Faunce
Fisheries Monitoring and Analysis Division, Alaska Fisheries Science Center, National Marine Fisheries Service, NOAA, Seattle,WA, craig.faunce@noaa.gov

Julie Bonney
Alaska Groundfish Data Bank, Kodiak, AK, jbonney@gci.net

Robert Swanson
Fisheries Monitoring and Analysis Division, Alaska Fisheries Science Center, National Marine Fisheries Service, NOAA, Kodiak Fisheries Research Center, Kodiak, AK

Fisheries Research 174 (2016) 219–233

Contents lists available at ScienceDirect

Fisheries Research

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

A field test of fisheries observer sampling methods for estimation of at-sea discards

Jennifer Cahalan[a,*], Craig Faunce[b], Julie Bonney[c], Robert Swanson[d]

[a] Pacific States Marine Fisheries Commission, 7600 Sand Point Way NE, Seattle, WA 98115, USA
[b] Fisheries Monitoring and Analysis Division, Alaska Fisheries Science Center, National Marine Fisheries Service, NOAA, 7600 Sand Point Way NE, Seattle, WA 98115, USA
[c] Alaska Groundfish Data Bank, P.O. Box 788, Kodiak, AK 99615, USA
[d] Fisheries Monitoring and Analysis Division, Alaska Fisheries Science Center, National Marine Fisheries Service, NOAA, Kodiak Fisheries Research Center, 301 Research Court, Kodiak, AK 99615, USA

ARTICLE INFO

Article history:
Received 30 October 2014
Received in revised form 3 October 2015
Accepted 3 October 2015

Keywords:
Groundfish fisheries
At-sea discards
Discard estimation
Gulf of Alaska
Sampling
Trawl catcher vessels

ABSTRACT

Estimation of at-sea discards is an issue that has received considerable attention worldwide. With this increasing focus, there is a need for greater precision of weight estimates for less common and fishery-limiting species. While one solution is to mandate full (100%) observer coverage to reduce uncertainty in estimation at the trip level, variance from on-deck sampling methods (e.g. within-haul) should also be properly addressed. Commercial fishing vessels are not perfect sampling platforms and all sampling methods suffer from implementation issues that potentially impact the quality of the data collected and the resulting estimates. We conducted a cooperative study with industry to evaluate two observer sampling methods on trawl vessels delivering their catch to shoreside processors. The alternative observer sampling method that targets the portion of the haul that would be discarded directly at-sea, relying on shoreside reports of retained catch to generate total catch estimates, was tested against the standard methods currently used by the NMFS North Pacific Groundfish and Halibut Observer Program that sample the entire catch, both retained and discarded portions (combined). Methods were tested simultaneously by deploying two observers to sample each haul on study trips within three Gulf of Alaska trawl fisheries that varied widely in amount and species composition of discards: Rockfish Program, arrowtooth flounder, and shallow water flatfish. Although the alternative method was successfully implemented in two of the three fisheries, logistical constraints decreased sampling effectiveness in the third. In some situations, observers were unable to collect multiple samples under both methods, preventing variance estimation. This occurred more often for the observer using standard methods. Detection of less common and rare species was higher using the alternative sampling method. Discard estimates from the two methods were found to be significantly different in two of the fisheries examined (Rockfish Program and arrowtooth flounder). Discard estimates under the alternative method tended to have smaller variances than for the standard method, although this was not universally the case. These results provide an important comparison of the relative performance of different on-deck sampling methods under varying catch conditions and fisheries.

1. Introduction

Estimating at-sea discards is an issue that has received considerable attention worldwide. Global and local fisheries catch including discards has been estimated (e.g., Alverson et al., 1996; Rochet et al., 2002; Kelleher, 2005), optimal sample sizes have been determined for programs that employ hierarchical designs (e.g., Allen et al., 2002), factors that are associated with discards have been identified (Feekings et al., 2012), and the best ways to estimate catch and discards have been explored (Tamsett and Janacek, 1999; Borges et al., 2004). Despite these important endeavors, we are aware of few studies that compare the perhaps more limited, but no less important, relative performance of on-deck sampling methods under varying catch conditions and fisheries (see, e.g., Tamsett et al., 1999a,b).

* Corresponding author.
E-mail addresses: jcahalan@psmfc.org (J. Cahalan), craig.faunce@noaa.gov (C. Faunce), jbonney@gci.net (J. Bonney).

http://dx.doi.org/10.1016/j.fishres.2015.10.004

The design of on-board sampling methods for observer programs needs to take into account the logistics of fishing operations (feasibility of implementation, see Tamsett et al., 1999a,b) as well as the management needs for data. As such, sampling methods used by observers in different programs will differ. For example, observers in the NMFS West Coast Observer Program sample the at-sea discards directly, usually weighing all discards (NWFSC, 2015). Retained catch is determined from industry landings data. However, haul-specific fishing locations are not available from these landings data. In contrast, in Alaska trawl fisheries, the volume of catches and discards often exceeds amounts that can be weighed directly by the observer. On many vessels, estimates of both retained catch and discarded catch are based on observer data, particularly on larger factory trawlers (Cahalan et al., 2014). While on vessels that deliver their catches to shoreside processors (catcher vessels, CVs), estimates of catch discarded at sea are based on observer data, and retained catches are based on industry-reported landings data. Observer sample methods in Alaska are focused on sampling the total catch and estimating the proportion of species-specific catch that is retained to be consistent in monitoring of fisheries across a wide range of vessel types and to provide haul-specific estimates of total catch size.

Monitoring of catch from the Alaska groundfish fisheries is the responsibility of the Alaska Regional Office Division of Sustainable Fisheries and the Alaska Fisheries Science Center's Fisheries Monitoring and Analysis Division (FMA). The FMA manages the North Pacific Groundfish and Halibut Observer Program (Observer Program) which deploys observers into commercial fisheries to collect catch data. In the North Pacific, there has been an increasing fishery management focus on individual vessel accountability of bycatch, and consequently there is an increasing need for higher precision of weight estimates for less common and fishery-limiting species discarded at sea, such as Pacific halibut (*Hippoglossus stenolepis*) and salmon (*Oncorhynchus* spp.).

In Alaska, non-pelagic (demersal) trawl catcher vessels (CVs) present one of the most challenging at-sea sampling situations that observers encounter (Cahalan, 2010). Individual hauls are generally 4.6–16.5 t, and crews sort catch directly from the trawl alley. On these vessels, observers work on the trawl deck where sampling space and access to catch is limited; the mean haul-level sample fraction in these situations is 6.6% of the total haul weight.

Many factors contribute to the variance of estimates of species-specific catches and at-sea discards, including difficulty in: (1) collecting samples and resulting small sample fractions, (2) estimating the weight of the total catches, and (3) determining the percentage of catch that is discarded at sea for each species. The haul-specific catch estimates may have high variance due to these sampling challenges. Sampling methods that are tailored specifically to this sampling environment may provide data for unbiased, lower variance estimation of catches.

Sampling efficiency on catcher vessels may be gained by using methods that target the portion of the total catch that is of interest, in this case at-sea discarded catches. Because the total weight of samples is limited by logistical constraints, the sample fraction achieved by observers collecting the same weight of sample from a smaller (e.g., discarded) portion of the catch is likely to be greater than that obtained using the standard Observer Program methodology (AFSC, 2011).

Our research compared two observer sampling methods in the trawl catcher vessel fleet: (1) the standard sampling method currently used by the Observer Program that samples from the entire catch (retained and discarded), and (2) an alternative method that is based on sampling the discarded portion of the catch thereby relying on industry-reported landings data of retained catch (at-sea methods similar to those used by the NMFS West Coast Observer Program; NWFSC, 2015). Specifically, we evaluated the potential gains in precision resulting from a more targeted sample method (sample discards only at the haul level; retained catch is based on landings data at trip level) against standard methods where total catch is sampled (both retained and discarded) at the haul level. Although the bias of the data resulting from these two sampling methods is not directly estimated because the actual weight of discards is not known, the relative accuracy of the two methods can be evaluated by comparing species-specific discard estimates for each haul. The objective of this research was to: (1) evaluate the feasibility of implementing this proposed alternative at-sea discard sampling method that targeted at-sea discards instead of total catches across a range of fisheries and fishing operations, (2) assess the potential for bias in the proposed sampling method relative to the standard method, and (3) evaluate the precision of estimated at-sea discard for these two methods.

2. Methods

2.1. Study fisheries

Participating vessels fished several trips in three fisheries; Gulf of Alaska rockfish (Rockfish Program), Arrowtooth flounder (*Atheresthes stomias*), and shallow water flatfish. The Rockfish Program is a catch share fishery with full observer coverage (observers onboard for every fishing trip), with catches dominated by rockfish species (*Sebastes* spp.) and generally small amounts of low diversity at-sea discard. The shallow water flatfish complex consists of northern rock sole (*Lepidopsetta polyxystra*), southern rock sole (*Lepidopsetta bilineata*), yellowfin sole (*Limanda aspera*), butter sole (*Isopsetta isolepis*), starry flounder (*Platichthys stellatus*), English sole (*Parophrys vetulus*), sand sole (*Psettichthys melanostictus*) and Alaska plaice (*Pleuronectes quadrituberculatus*). The arrowtooth flounder and shallow water flatfish fisheries have observers onboard for some of their fishing trips. Discards in the latter two fisheries are more diverse and the total amount of at-sea discard is much higher than seen in the Rockfish Program.

2.2. At-sea sampling methods

Two observers were deployed on each trip and hauls were sampled simultaneously by both observers: one observer sampled using the standard sampling method in which the total (unsorted) catch is sampled while the other observer used the alternative sampling method to only sample the at-sea discards (paired analysis). In this study we compared the precision and relative bias of the two methods. The precision of estimates was assessed for the two methods by comparing the estimated variances of the discards. Bias of the two estimates based on data from the two methods was evaluated in a relative sense; the species-specific catch was not weighed and hence the true value was not known.

The major difference between the standard and alternative sampling methods is in the definition of the sampled population within each haul. The standard Observer Program method for at-sea sampling in Alaska defines the sampled population as the *entire* catch (retained and discarded), while the alternative method defines the sampled population as the *portion of the catch that was discarded*. The objective of sampling in both situations is to collect data to estimate the at-sea discard of catch for each sampled haul, as well as include recording haul-specific catch locations (latitude and longitude), collecting biological data (e.g. fish lengths, otoliths, and halibut injury data), monitoring for marine mammals and seabirds, including recording and monitoring gear interaction with US Endangered Species Act species.

The standard Observer Program protocols were used to select which hauls on a trip were sampled; that is, both observers sampled

as many hauls as logistically possible or randomly selected hauls to be sampled. Which of the observers used the alternative method and which used the standard method was randomly determined on the first haul of a trip; each observer alternated the sampling method they used thereafter.

2.2.1. Standard sampling methods

The standard sampling method is a simple random sample taken from the *total* catch of each haul. The observer assessed each haul as it was brought on board noting the species diversity of the catch, the vessel operations and crew sorting behaviors, and available deck space. For each sampled haul, the observer determined the total volume of the catch by measuring the cod end of the net in several places and applying a volume formula appropriate for the general shape of the codend. The density of the catch was estimated from a sample of known volume and weight of a sample of unsorted catch. The observer estimate of total weight of catch was estimated as the product of the cod end volume and the catch density.

The observer determined how to best randomize their sampling of the catch for species composition for each sampled haul by dividing the catch for the haul into defined sampling units, determining number of units they are able to collect, and deciding how to randomize the collection of those samples (Cahalan et al., 2014; AFSC, 2011). Sample units are typically defined in terms of catch weight or volume, e.g. a fixed weight of fish (100 kg) or a container of known volume (tote or basket). The observers randomly or, in the case that randomization was not possible, opportunistically, selected sample units for data collection and obtained the weight of each species in the sample and the total weight in the sample.

The observer recorded an estimate of the proportion of each species that was retained and discarded based on the operational characteristics of the vessel and crew, visual monitoring of discard activities, information about discards collected from the captain and crew and on occasion, direct measurement of weight of discards. On trawl vessels, the proportion of each species retained is generally not based on sample data. Additional sampling method details can be found in AFSC (2011).

Estimates of discard were generated using the analytical methods described in the next section. In brief, the species proportion of the catch in the samples was expanded to the haul by multiplying the species proportion (from the sample data) and the total (unsorted) haul weight. This total estimated catch weight for each species was multiplied by the visually estimated proportion discarded (1-proportion retained) to generate weight estimates of the at-sea discards for each species.

2.2.2. Alternative sampling methods

The alternative sampling method is a stratified random sample collected from the portion of the catch that will be discarded at sea. Observers defined strata of at-sea discards based on the sorting practices of the crew. When some species were sorted into categories of catch (e.g., Pacific halibut), these were defined as a separate stratum and sampled independently of other sorted discard catch.

Within each stratum, the to-be-discarded catch was divided into equal-sized sample units, typically based on either time or volume depending on vessel configuration and crew activities. The observer then randomly selected one or more sample units from the discard stratum and weighed each species in the sample. For example, if the catch was sorted into bins on deck, the observer may have defined sample units in terms of bins or baskets and randomly selected baskets of discard to sample. However, the observer may have defined sample units in terms of time and randomly selected time periods during which all discards were collected as part of the sample if a conveyor belt system (sorting belt) was used in sorting the catch from the trawl alley. If possible, the observer weighed the entire catch of each species in the discard (census of discards).

The species proportion of the catch in the samples was expanded to the size of the haul using the analytical methods below. Since the observer was sampling directly from the at-sea discards, estimates of species-specific discard weight were generated. However, total catch estimates and estimates of retained catch were not available from data collected using the alternative method. Additional details on this method are in given in Faunce et al. (2013).

2.2.3. At-sea methods pre-test

A test of at-sea sampling methods was conducted in October 2010 to assess feasibility of planned study methods. Two study biologists were deployed for a 4-day trip on a catcher vessel (CV) trawler in the arrowtooth flounder fishery during which they tested the proposed sample methods. All six hauls that were fished on that trip were sampled and the entire catch from one haul was weighed. No difficulties were identified during that trip, and hence no substantive changes were made to sampling methods and the results from that trip were included in the final analyses.

2.2.4. At-sea study implementation

The following spring and summer, eight additional CVs using trawls agreed to participate in the study. Nine observers were deployed in pairs on 12 trips during which 2–15 hauls were sampled per trip (Table 1). Fishing occurred in three fisheries with different fishing and discarding characteristics. Of the 94 hauls that were sampled, neither sampling method detected discards on three hauls in the Rockfish Program fishery.

Three hauls were sampled by only one observer, two in the shallow water flatfish fishery and one in the arrowtooth flounder fishery. With the exception of one vessel, all participating vessels had sorting belts although they were not consistently used throughout the study. In the Kodiak trawl catcher vessel fleet, most vessels have sorting belts; hence this type of operation is consistent with the larger population of vessels.

2.3. Analysis methods: at-sea catch and discard estimation

Estimates of total catch were generated for each species by expanding the species composition from the observer sample to the haul consistent with the sampling method used to collect the data. Using the standard sampling method, this sample fraction is the total weight of the sample divided by the total weight of the haul, and the at-sea discard is determined by multiplying the catch of each species by the estimated discard rate. However, the sample fraction is the number of sample units selected divided by the total number of sample units in the discard population using the alternative sampling method. The sample frames of these two methods are defined differently and hence the estimation methods used to extrapolate catches to the total discards for the haul are different.

2.3.1. Discard estimates and variance components: standard method

Observers using the standard sampling method used a simple random sample design to sample from the total catch. All the estimators described in this section assume this simple random selection of samples. In some cases, observers used a systematic random sample design. The assumption of a simple random sample when a systematic sample is taken will tend to result in overestimation of the estimated variance (Thompson, 1992). In other cases, observers were not able to obtain randomly selected samples and instead selected opportunistic samples. These are also treated as simple random samples.

The total haul weight is computed as the product of the volume of the haul (cod end or bin volumes) and the density of the haul. The

Table 1
Sampling logistics summary.

	Rockfish Program	Arrowtooth Flounder	Shallow water flatfish
Data collected	3 vessels, 3 trips, 28 hauls	7 vessels, 7 trips, 41 hauls	3 vessels, 3 trips, 25 hauls
Description	Catch share program; 100% retention of quota species	Limited access fishery	Limited access fishery
Total catch	8–10 t	6–18 t	3–8 t
At-sea discards	Under 1 t (<5%)	1–3 t (5–40%)	2–5 t (>50%)
Diversity of discards	1–6 species	3–22 species	5–20 species
Alternative method impact to fishing operations	None	Increased crew workload	Increased crew workload
Alternative method tracking of discard activity	Effective	Somewhat effective	Not effective
Effectiveness of implementation	Able to weigh all discards: *need single observer*	Able to sample all discards; inconsistent ability to obtain multiple random samples: *need two observers*	Unable to sample all discards; low ability to obtain multiple random samples: *need two observers*

haul weight, $\hat{W}$, is estimated using Eq. (1), where w_j is the weight of fish in sample j ($j = 1, \ldots, J$ samples), v_j is the volume of sample j, and V is the volume of the haul. Variance estimates are not available for the total haul weight.

$$\hat{W} = V\left(\frac{\sum_{j=1}^{J} w_j}{\sum_{j=1}^{J} v_j}\right) \quad (1)$$

The estimator used to generate haul-specific catch is given by Eq. (2), below, where $\hat{W}$ is the weight of the haul, d_i is the proportion discard of species i and w_{ij} is the weight of species i in sample j. This is nominally a ratio estimator that assumes that the total haul weight ($\hat{W}$) is a known constant. In addition, since the proportion discard (d_i) is based on direct observation (not sample data), it is also assumed known without variance. As above, the haul subscript has been omitted for clarity; all estimates are haul and species (i) specific. The subscript S denotes the standard sampling methodology was used.

$$\hat{W}_{iS} = \hat{W}\frac{\sum_{j=1}^{J} d_i w_{ij}}{\sum_{j=1}^{J} w_j} = \hat{W}\frac{d_i \bar{w}_i}{\bar{w}} \quad (2)$$

For this analysis, we assume that there is no correlation between $\bar{w}_i$ and $\bar{w}$; the size of the sample is not a good predictor of the species-specific weight in the sample, hence this is not a true ratio estimator. The weight of any given species cannot be greater than the total weight of the sample, so $\bar{w}_i$ and $\bar{w}$ are not completely independent. However, where the species of interest is not common in the catch and/or occurs in the haul in a clustered fashion, the weight of the sample and the weight of the species in the sample may be close to independent, and the covariance term generally included in the ratio-estimator will be (close to) zero.

An approximate variance derived from a first-order Taylor expansion of two independent variables was used to estimate $V\hat{a}r(\hat{W}_{iS})$, Eq. (3), where p_w is the sample fraction based on sampled weight. Both $V\hat{a}r(\bar{w}_i)$ and $V\hat{a}r(\bar{w})$ were estimated using the usual sum of squares formulation. Note that the finite population correction factor $(N-n)/N = (1-p_w)$ (Thompson, 1992) is included in the variance estimator.

$$V\hat{a}r(\hat{W}_{iS}) = W^2 d_i^2 (1-p_w)\left[\bar{w}_i^2 \bar{w}^{-4} V\hat{a}r(\bar{w}) + \bar{w}^{-2} V\hat{a}r(\bar{w}_i)\right] \quad (3)$$

2.3.2. Estimates and variance components: alternative method

The observer who sampled using the alternative method used a stratified sampling design to sample only the portion of the catch that would be discarded. Strata were defined by crew sorting activities. Crew sorted catch to be discarded into bins on-deck that contained the catch that would be discarded until the observer could sample; once the observer sampled, or the bin was completely filled, the catch was discarded at sea. Each bin was sampled as an independent stratum. The observer sampled each stratum, recording the number of sample units in the stratum, N, and in the sample, n. This departure from standard methods allows design-based estimation based on expansion of the sample mean to the stratum, instead of relying on ratio-type estimators described above. If there are two or more discard strata, the estimates for each stratum are combined to a single discard estimate for each species in the haul, Eq. (4), where h indexes the stratum and other subscripts are as defined previously. The variance is computed using the usual estimator for a stratified simple random sample (Thompson, 1992; Eq. (5)).

$$\hat{W}_{iA} = \sum_{h=1}^{H} N_h \bar{w}_{hi} \quad (4)$$

$$V\hat{a}r(\hat{W}_{iA}) = \sum_{h=1}^{H} N_h^2\left(1-\frac{n_h}{N_h}\right)V\hat{a}r(\bar{w}_{hi}) = \sum_{h=1}^{H} N_h^2\left(1-\frac{n_h}{N_h}\right)\frac{\sum_{j=1}^{J_h}(w_{ij}-\bar{w}_i)}{J_h(J_h-1)} \quad (5)$$

Since observers were not always able to obtain samples from each of the strata, the assumption was made that the strata that were not sampled were similar in composition to those that were sampled. The total estimated weight and variance of weight was generated by expanding the estimates from the sampled strata to the size of the entire discarded catch (all strata, sampled and not), equations 6 and 7, where h_s indexes the sampled strata.

$$\hat{W}_{iA} = \frac{\sum_{h=1}^{H} N_h}{\sum_{h_s=1}^{H_s} N_{h_s}} \sum_{h_s=1}^{H_s} N_{h_s} \bar{w}_{h_s i} \quad (6)$$

$$V\hat{a}r(\hat{W}_{iA}) = \left[\frac{\sum_{h=1}^{H} N_h}{\sum_{h_s=1}^{H_s} N_{h_s}}\right]^2 \sum_{h_s=1}^{H_s} N_{h_s}^2\left(1-\frac{n_{h_s}}{N_{h_s}}\right)V\hat{a}r(\bar{w}_{h_s i}) \quad (7)$$

Variance estimates are not available for hauls where the observers were not able to obtain multiple samples in each sampled

strata. Data from these strata were used in estimation of discards. However, estimates from hauls where all strata did not have multiple samples were not included in comparisons of precision (since no variance estimates were available).

2.3.3. Species detection

A logistic regression was fit to the data to evaluate detection rates of rare species in the at-sea discards between the two methods. Standard logistic regression methods were used (e.g., McCulloch and Nelder, 1989) and analysis was conducted using R (R Development Core Team, 2014). Model fit was based on a likelihood ratio test (McCulloch and Nelder, 1989). A species was considered 'detected' by a methodology if that species was present in sample data for that sample method and 'not detected' if that species was not present in the sample data for that method, but was present in the other sample method data.

Covariates in the detection analysis included sampling method, fishery and discard-species proportion. Species proportion was defined as the species-specific weight of the discard divided by the total haul weight; if the species was detected in the sample data from both methods, the average species proportion was estimated as the mean of the estimates from the two methods while if the species was detected by one method and not the other, the single observation of species proportion was used. By using the average of the species proportion of the discard from both methods, the proportion was not dependent on a single sampling method, and provides an indication of the species prevalence in the discarded catch.

2.3.4. Differences in discard estimates

In order to assess whether data collected using both methods were generating similar discard estimates for each haul, the haul-specific difference in species-specific weight was computed, Eq. (8), where as previously i indexes species and the haul subscript is omitted.

$$\hat{D}_i = \hat{W}_{iS} - \hat{W}_{iA} \tag{8}$$

Since the data are collected using two sampling methods on each haul, the estimates are independent and the variance of the difference is the sum of the variances estimated for each discard methodology: $Var(\hat{D}_i) = V\hat{a}r(\hat{W}_{iS}) + V\hat{a}r(\hat{W}_{iA})$. We want to test the null hypothesis $H_o : \hat{D}_i = 0$, however, since the magnitude of $\hat{D}_i$ varies with each species, trip, and fishery, $\hat{D}_i$ was standardized to the sum of the two observations, which is the relative difference between the two observations, Eq. (9). We used the sum of the estimates since neither estimate was assumed to be more correct than the other estimate, and the true value was unknown.

$$\hat{R}_i = \frac{\hat{D}_i}{\left(\hat{W}_{iS} + \hat{W}_{iA}\right)} = \frac{\hat{W}_{iS} - \hat{W}_{iA}}{\left(\hat{W}_{iS} + \hat{W}_{iA}\right)} \tag{9}$$

Wilcoxon tests were used to test whether the difference in median relative discard differences ($\hat{R}_i$) was equal to zero while generalized linear models were used to test whether the mean of relative discard differences was equal to zero and what factors contributed to any observed differences (e.g. fishery). These analyses assumed that the $\hat{R}_i$ were measured without variance (e.g., the haul level discard estimates are without variance, i.e., $V\hat{a}r_i(R_i) = 0$).

However, noting that the $\hat{R}_i$are themselves estimates that are distributed with some mean and variance, the variance of $\hat{R}_i$ is derived using a first order Taylor expansion of a ratio; since x and y are independent, the variance of $(x+y)$ and the variance of $(x-y)$ are the same, and the covariance of $(x+y)$ and $(x-y)$ is equal to $V\hat{a}r(x) - V\hat{a}r(y)$, Eq. (10).

$$V\hat{a}r(R_i) = V\hat{a}r\left(\frac{\hat{W}_{iS} - \hat{W}_{iA}}{\hat{W}_{iS} + \hat{W}_{iA}}\right) = \frac{\hat{W}_{iS} - \hat{W}_{iA}{}^{2}\, V\hat{a}r\hat{W}_{iS} + \hat{W}_{iA}}{\hat{W}_{iS} + \hat{W}_{iA}{}^{4}} + \frac{V\hat{a}r\hat{W}_{iS} - \hat{W}_{iA}}{\hat{W}_{iS} + \hat{W}_{iA}{}^{2}} - \frac{2\hat{W}_{iS} - \hat{W}_{iA} V\hat{a}r\hat{W}_{iS} - V\hat{a}r\hat{W}_{iA}}{\hat{W}_{iS} + \hat{W}_{iA}{}^{3}}$$

$$V\hat{a}r(R_i) = \frac{\hat{D}_i^{2}\, V\hat{a}r\hat{D}_i}{\hat{W}_{iS} + \hat{W}_{iA}{}^{4}} + \frac{V\hat{a}r\hat{D}_i}{\hat{W}_{iS} + \hat{W}_{iA}{}^{2}} - \frac{2\hat{D}_i(V\hat{a}r\hat{W}_{iS} - V\hat{a}r\hat{W}_{iA})}{\hat{W}_{iS} + \hat{W}_{iA}{}^{3}} \tag{10}$$

To incorporate this variability into the analysis, the $\hat{R}_i$ were further standardized by assuming that each $\hat{R}_i$ is distributed with mean zero (under null hypothesis) and some variance $V\hat{a}r(\hat{R}_i)$. Then using the central limit theorem and Eq. (10), we constructed a set of standardized relative differences, $\hat{S}_i$, each distributed normally with mean of 0, and variance equal to 1 (standard normal distribution), using Eq. (11).

$$\hat{S}_i = \frac{(\hat{R}_i - 0)}{\sqrt{V}} \tag{11}$$

If the standardized relative discard differences $\hat{S}_i$ are each from a standard normal distribution, $N(0, 1)$, we would expect only 5% of these observations to fall outside of the 2.5th and 97.5th percentiles of the standard normal curve. If the two sampling methods yielded different discard estimates, then we would not expect that the set of $\hat{S}_i$ to be distributed according to a standard normal; in particular, we would expect more than 5% of the $\hat{S}_i$ to be outside the 2.5th and 97.5th percentiles (i.e., farther from zero).

2.3.5. Differences in variance estimates

The variance estimates were compared to test whether the variance of the discard estimates is the same using the two methods. The estimated variance is considered known for each species and haul (i.e., does not itself have variance). Since the magnitude of the variance will be a function of the discard estimate, the percent standard error (PSE, also referred to as relative standard error, RSE, Jessen 1978) of the estimates was used in the analysis. The PSE for each observation (haul, species, and method) was computed in the usual manner as the square root of the variance of the estimate divided by the discard estimate, Eq. (12), where k is indexing the sampling method and the haul subscript is omitted for clarity.

$$PSE_{ik} = 100\frac{\sqrt{V\hat{a}r(\hat{W}_{ik})}}{\hat{W}_{ik}} \tag{12}$$

The difference in the PSE was computed as the standard PSE minus the alternative PSE; $P_i = PSE_{iS} - PSE_{iA}$. This difference was used as the response variable in Wilcoxon tests to assess whether the median PSE difference was equal to zero. In addition, generalized linear models (McCullagh and Nelder, 1989) were used to test whether mean difference varied with fishery and species discard proportion.

3. Results

3.1. Summary of sampling results

The amount of at-sea discarded catch differed by fishery with almost no discards in the Rockfish Program fishery and more discards in the shallow water flatfish fishery (Fig.1, right panel). The

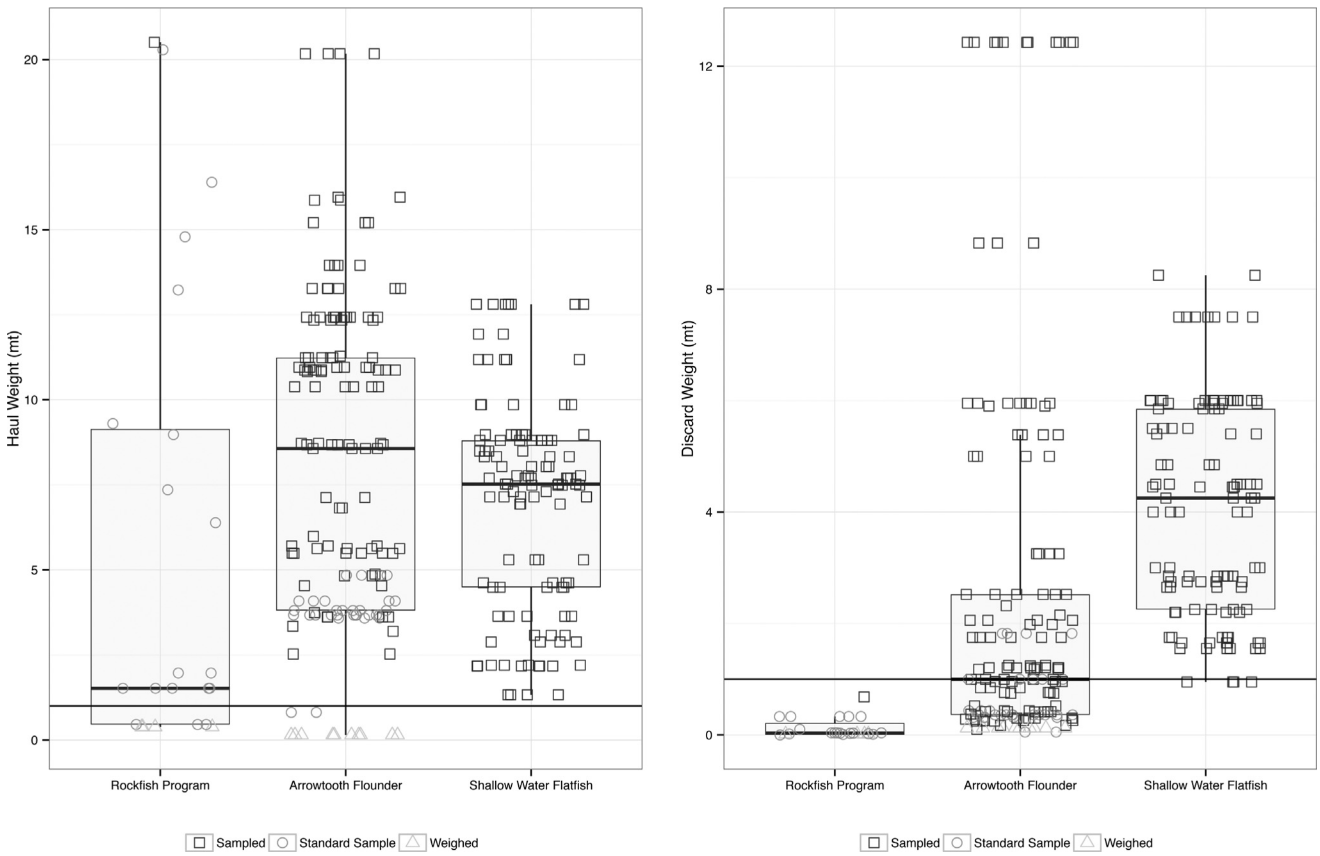

Fig. 1. Haul size (left panel) and at-sea discard amount (right panel) by each fishery. Individual data points are either the observer estimates, or when two estimates are available, the mean estimate. Shape of data points indicates how data were collected: both observers sampled (sampled), standard observer sampled and alternative observer weighed all discards (standard sample), or both observers weighed all discards (weighed).

median size of the hauls in the Rockfish Program fishery was smaller than in the arrowtooth flounder and shallow water flatfish fisheries (Table 1, Fig. 1, left panel). Haul sizes varied considerably, ranging from less than 5 t to more than 20 t, with most hauls weighing between 5 t and 15 t based on the observers' estimates of total catch.

A summary of haul-specific sampling results (i.e., numbers of samples, sample fractions) is presented in Table 2. The observers using the standard method were able to completely weigh the *total catch* of two hauls (arrowtooth flounder fishery and Rockfish Program fishery) while observers using the alternative method were able to completely weigh all *discarded catch* on 33 hauls. On sampled (not completely weighed) hauls, standard method sample fractions ranged from approximately 1% to 60% (mean of 8%) *of the total catch* in the Rockfish Program fishery, 1–30% (mean of 6%) in the arrowtooth flounder fishery, and 1–20% (mean of 6%) in the shallow water flatfish fishery. For hauls where the alternative method observer did not weigh all the discarded catch, the alternative method sample fractions were 1–98% (mean of 23%) *of the discards* in the arrowtooth flounder fishery, 5–40% (mean of 20%) in the shallow water flatfish fishery and 30% for the single sampled (i.e., not a census) haul in the Rockfish Program fishery.

The standard method observer generally collected a single sample per haul in the Rockfish Program and arrowtooth flounder fisheries. However, they collected three samples per haul in the shallow water flatfish fishery (Table 2). The alternative method observers in the Rockfish Program fishery collected three samples from the single sampled stratum in that haul. However, in the shallow water flatfish fishery there were ten hauls where three or more samples were taken from the sampled stratum and 15 hauls where only one sample was taken in each of 55 of the strata. In this fishery only, the observer was not able to sample all strata (eight hauls). For the 35 hauls with sampled strata in the arrowtooth flounder fishery, 20 strata were sampled with three or more samples each while six strata had only two samples per strata.

3.2. Feasibility of alternative sampling method

The observer's ability to sample the at-sea discards using the alternative method varied with fishery. Based on feedback from the project observers, the sampling difficulties increased with the amount of discard encountered and tended to be related to tracking crew discarding and sorting behavior, insufficient time to collect samples or space to store discards, and safety concerns. This impacted the quality of the data, particularly in the arrowtooth flounder and shallow water flatfish fisheries.

The vessel crews typically did not use the sorting belts in the Rockfish Program fishery. The observers on one vessel were not able to obtain independent measures of the cod end volume for most hauls since vessel operations did not allow sufficient time for measurements to be taken before dumping the codend onto the trawl deck. In these cases, the vessel estimate (hail-weight) of total catch was used to expand the species composition from the samples to the haul. With the exception of one tow per trip that was predominantly Pacific cod (*Gadus macrocephalus*), in this fishery there were few discards and observers reported that the alternative method could be implemented with few problems (Fig. 1). In some cases, the only at-sea discard were the fish in the standard method samples.

Table 2
Haul sampling summary.

	Rockfish Program	Arrowtooth Flounder	Shallow water flatfish
Standard method			
Number samples per haul	1–3; 24 with single sample; 1 haul completely weighed; 15 hauls no discards	1–3; 23 with single sample; 1 haul completely weighed	3
Number hauls with 3 or more samples	1	5	25
Fraction of total catch sampled (non-censused hauls), range (mean)	1– 60% (8%)	1–30% (6%)	1–20% (6%)
Alternative method			
Number populations per haul	27 hauls completely weighed, 1 haul with 2 populations	6 hauls completely weighed, 35 hauls sampled, 1–7 populations per haul	25 hauls sampled, 1–9 populations per haul, 8 hauls with un-sampled populations
Number of hauls where all populations have 3 or more samples	1	20	10
Fraction of discards sampled (non-censused populations), range (mean)	30% (30%)	1–98% (24%)	5–40% (20%)

Note that the sample fraction from the alternative method is relative to the discarded catch while the sample fraction from the standard method is relative to the total catch.

The participating vessels in the shallow water flatfish and arrowtooth flounder fisheries tended to use their sorting belts for at least part of the haul processing. Observers on two vessels in the arrowtooth flounder fishery reported cases where there were large volumes of discards and the sorting belt had to be slowed or stopped during the sample period for the samples to be collected. Since a temporal frame is used, this may have resulted in underestimation of the amount of discard in the shallow water flatfish and arrowtooth flounder fisheries, although it is not possible to quantify the degree of bias caused by this activity.

Observers also reported that sorting belts had difficulty lifting large organisms. In one case, the observers segregated these organisms into a separate stratum while on another vessel these organisms were assisted up the belt by either the observer or a crew member. One observer noted crew sometimes threw larger species onto the sorting belt while the observer was sampling, possibly introducing a bias. The crew had opportunity to sort fish away from the sorting belt without the observer's knowledge. While this activity was only recorded by observers in the arrowtooth flounder fishery, similar situations were described by observers in the shallow water flatfish fishery.

Observers had the greatest difficulty implementing the alternative method successfully in the shallow water flatfish fishery. The volume and diversity of discards was typically high (Fig. 1, right panel). Observers were unable to collect multiple samples and were unable to monitor crew discards for all hauls. On all three vessels in this fishery, observers noted cases where crew sorted catch per usual vessel operations and the catch would have been discarded without the alternative observer's knowledge if the other observer had not been present to redirect the discard. On two vessels, observers sampled from the sorting belt while on the third vessel observers sampled from the trawl deck.

The observers in the shallow water flatfish and arrowtooth flounder fisheries noted that the alternative sampling method was much more physically demanding for both the observers and the crew. Vessel crew needed to handle catch twice; once to sort discards to specific locations for the observer to sample and a second time to discard them over board. In addition, they reported several safety concerns including traversing the trawl deck while carrying sampling gear and sampling in areas that exposed them to wave action. In cases where weather was poor, the observers anticipated that it would be problematic to store discards so that they could be sampled due to wave wash and deck motion.

3.3. At-sea discard estimation

Estimates of at-sea discards were generated from data collected using both sampling methods on 91 hauls where at-sea discards occurred. At-sea data were used to generate 1522 discard estimates; one estimate for each species on each haul sampled under each sampling method (Fig. 2). These were predominantly from the arrowtooth flounder (844 estimates; 56%) and shallow water flatfish trips (534 estimates, 35%), with the remaining 144 estimates (9%) from Rockfish Program trips.

There were fewer discards in the Rockfish Program fishery (Fig. 2). The percent standard error (PSE) for each discard estimate that was based on at least three samples was also computed and plotted around each of the catch estimates. Hauls where both observers were able to weigh the entire species-specific discards occurred only in the arrowtooth flounder and Rockfish Program fisheries; neither observer was able to obtain a census in the shallow water flatfish fishery (Fig. 2).

3.4. Species detection

Detection of a species in the discards was a function of both the sampling method and the prevalence of the species in the catch (Fig. 3). While species were detected using either method at low species proportions, failure to detect a species only occurred when the species was rare, <7% of the total catch. Of the at-sea discard estimates generated, 37.8% were on hauls in which the species was detected using both sampling methods.

The factors included in the logistic regression model were average discard species proportion, sampling method, and fishery. The average discard species proportion was the discard species proportion averaged across sample data collected using the two sample methods. If data from one sampling method did not include a particular species, the discard species proportion was based on data collected using the other method. Total haul weight and species type (flatfish, rockfish, invertebrate etc.) were tested using a likelihood ratio test, but were not significant factors and did not contribute to overall model fit; hence they were not included in the final model. The overall model fit was not rejected (deviance of 1533.6 distributed chi-square on 1517 df, p-value = 0.337) and the model was significantly better than the null model (decrease in deviance of 352.9 on 4 df, p-value «0.0001).

Model parameters (log-odds) are presented in Table 3. The odds of detecting a species using the alternative methodology are more than 4 times greater, all other factors held constant, than detecting the same species using the standard method. Similarly the odds of detecting a species are smaller in both the arrowtooth and shallow water flatfish fisheries relative to the Rockfish Program fishery, all other factors held constant. Both the species proportion and the sampling method were significant factors in determining the detectability of a species.

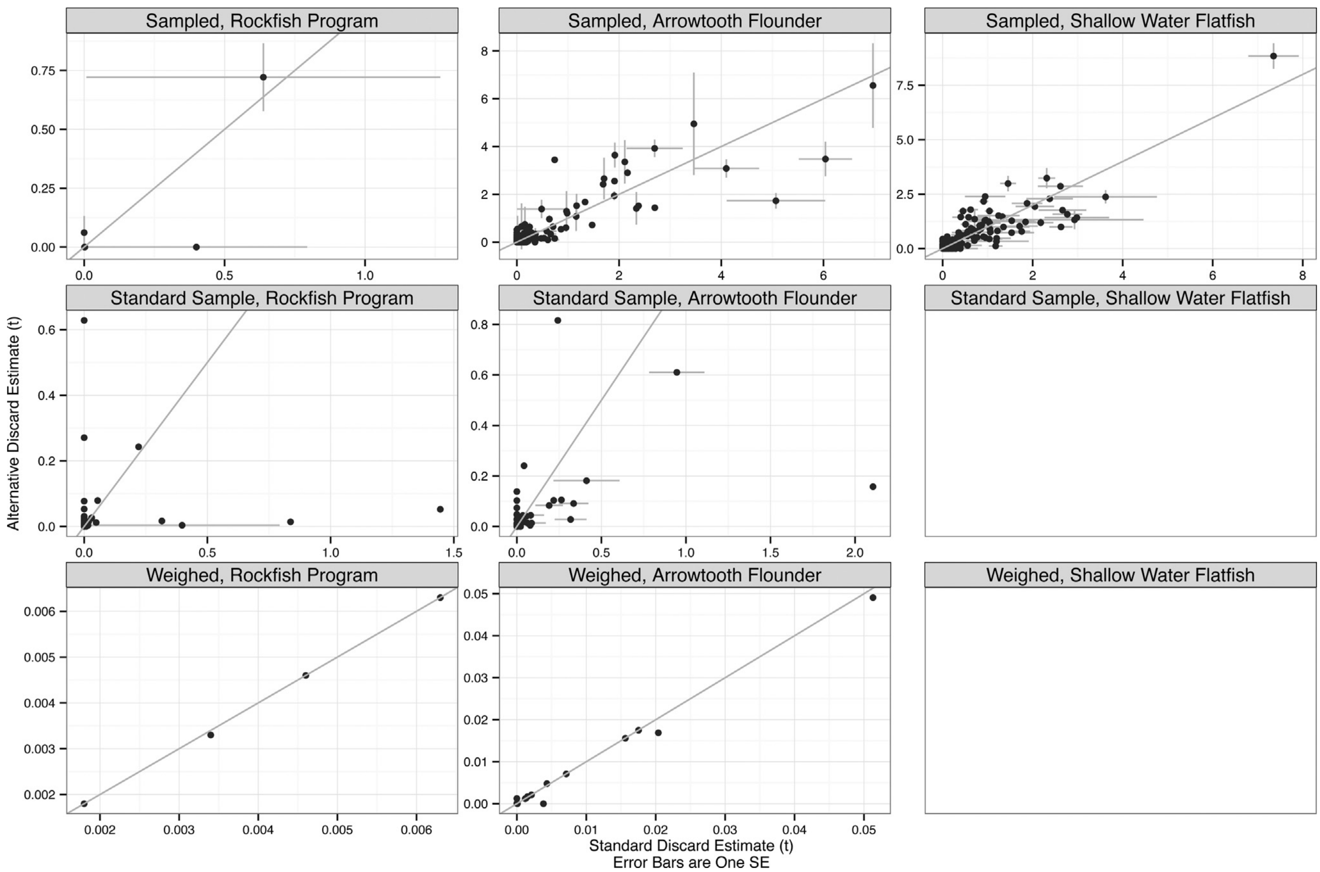

Fig. 2. Comparison of discard estimates generated using alternative sampling methods (y-axis) and standard methods (x-axis) data for each fishery (columns: arrowtooth flounder, Rockfish Program, and shallow flatfish). Solid line is the 1-to-1 reference line (i.e., alternative method estimate equal to standard method estimate). Error bars are one standard error for those cases where the discards were sampled and three or more samples were collected. Estimates generated from data where both observers sampled the discards are presented in the top row (neither was able to obtain a census). Estimated discard from data where only the observer using the standard method sampled and while the observer using the alternative method completely weighed the catch are presented in the middle row. The lower row of plots presents estimates based on both observers completely weighing all discards.

Table 3
Detection model coefficients and results. The standard method is used as the reference (baseline) in computing the odds ratio; the Rockfish Program fishery is also used as a reference level.

Parameter	Estimate	Standard error of estimate	*Z*-score (*P*-value)	Odds ratio	95% CI (on odds ratio)
Intercept	−0.25	0.198	−1.28 (0.20)		
Species proportion	44.90	5.323	8.44 («0.0001)		
Alternative method	1.42	0.127	11.24 («0.0001)	4.155	3.248, 5.340
Arrowtooth Flounder	−0.07	0.206	−0.36 (0.720)	0.929	0.617, 1.387
Shallow water flatfish	−0.30	0.221	−1.38 (0.168)	0.738	0.476, 1.133

The effect of species proportion and sample method can be seen in the plot of the predicted values over the range of data included in the model (Fig. 4). There is a 100% probability that a species is detected in the discards using either method and in all fisheries when the species-specific discard is greater than approximately 15% of the haul weight. As reported above, the alternative method has greater detection of discarded species than the standard method as the species proportion decreases.

3.5. Comparison of catch estimates

Discard estimates were compared for each sampling method, haul, and species. There were 288 relative difference-in-discard estimates for hauls where estimates for the species were made using data from for both sampling methods (species was detected by both observers), computed using Eq. (11). The null hypothesis of no difference in discard estimates from the two methods equates to a relative difference of zero. However, the range and median of the relative differences varied by fishery (Fig. 5). Although there were fewer observations in the Rockfish Program fishery, those relative differences tended to be greater than zero (standard method estimate larger) while for the other two fisheries the relative differences were more evenly distributed around zero.

Wilcoxon tests indicated that the relative differences in the medians were not equal to zero (Wilcoxon $V = 24904$, p-value = 0.00115). This pattern was not consistent for all three fisheries however. Generalized linear model methods were used to determine whether the relative differences varied significantly with fishery (Table 4). Shallow water flatfish was the only fishery where the mean relative discard difference was not significantly different from zero. This is also the only fishery where both the standard method and alternative method observer sampled at least part of every sampled haul, instead of directly weighing all or part of the discards.

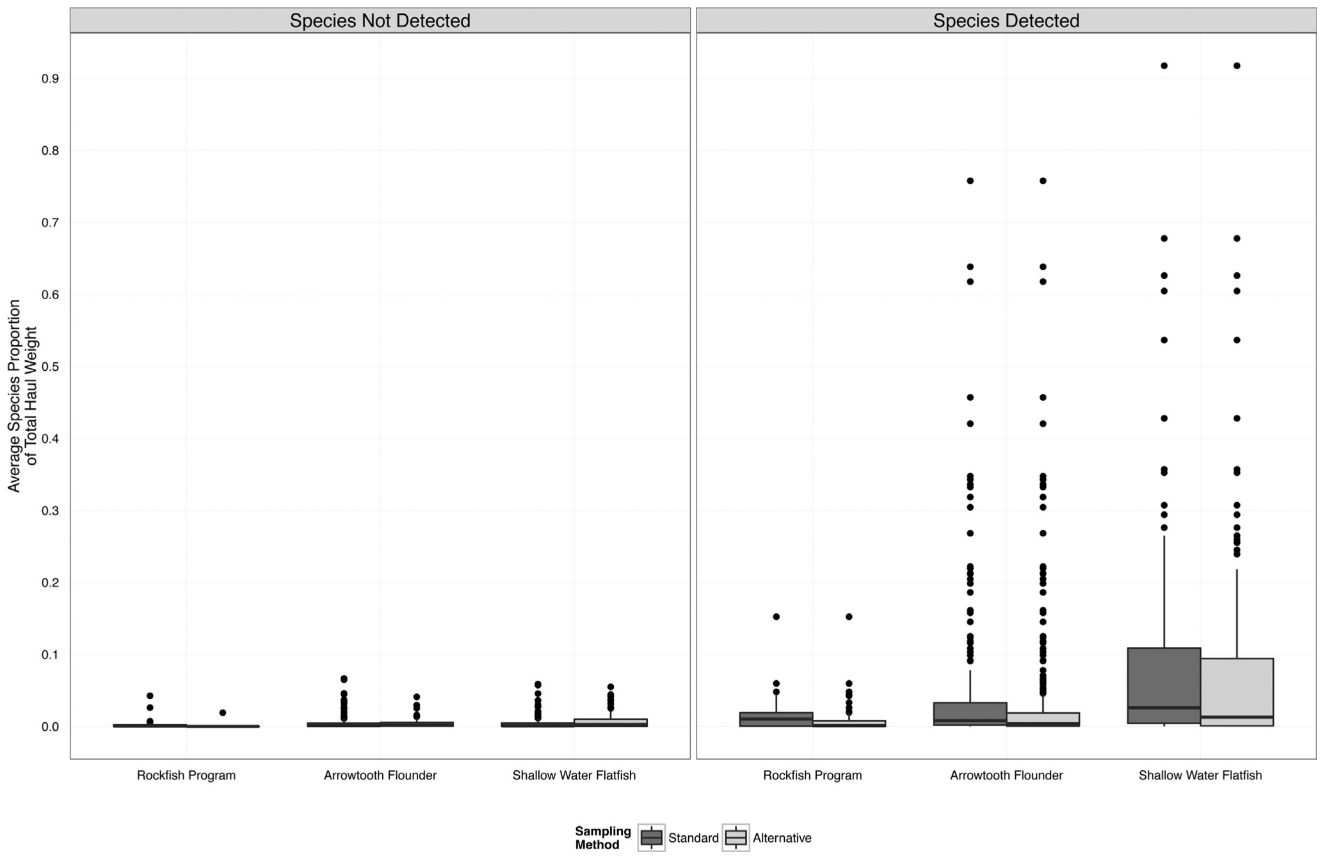

Fig. 3. Species detection as a function of species proportion of the catch (y-axis) and sampling method for each fishery (y-axis).

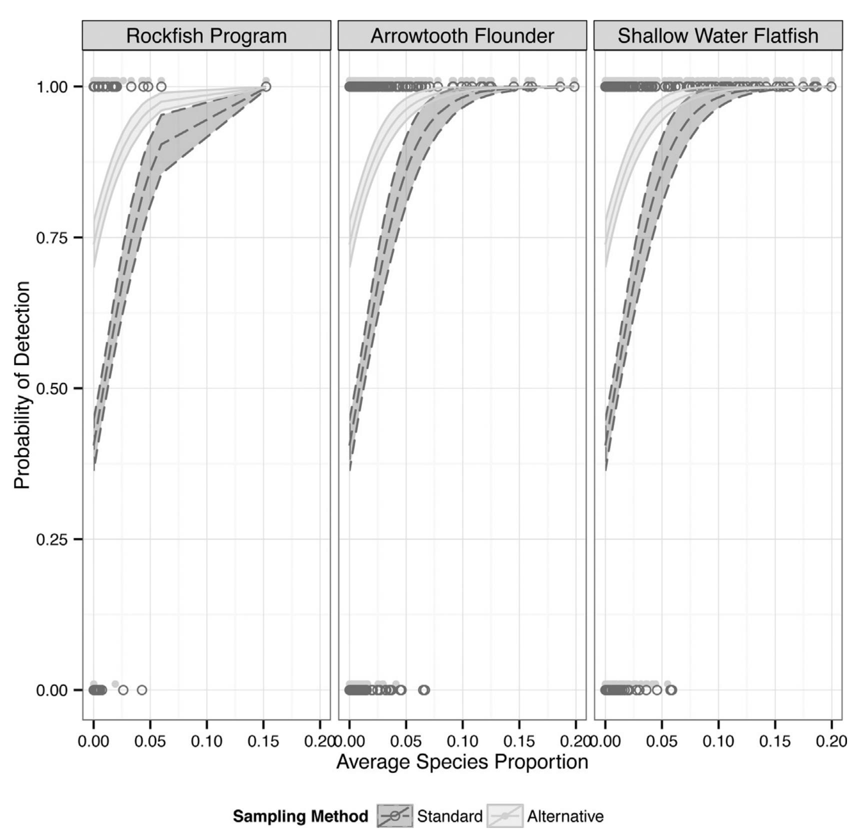

Fig. 4. Predicted probability of species detection based on logistic regression model presented in Table 3. Shaded areas indicate the 95% confidence intervals for the standard method (dashed line) and alternative method (solid line); individual data points are for both the standard method (open circle) and alternative method (closed circle).

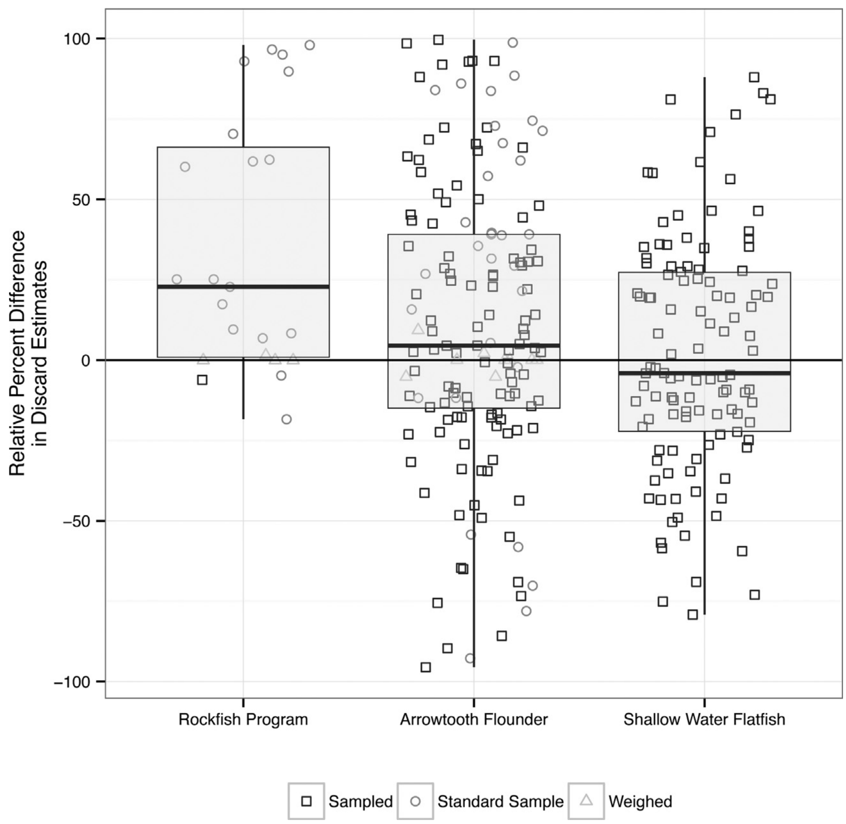

Fig. 5. Distribution of the relative difference in discard estimates for species that were detected by both sampling methods. Values less than 1 indicate the alternative method estimate was larger than the standard method estimate. Shape of data points indicates how data were collected: both observer sampled (sampled), standard observer sampled and alternative observer weighed all discards (Standard Sample), or both observers weighed all discards (Weighed).

Table 4
Relative discard estimate difference model results.

	Coefficient (standard error)	95% CI	*p*-Value (t)
Rockfish program	0.35 (0.086)	0.185, 0.523	«0.0001 (4.1)
Arrowtooth Flounder	0.11 (0.034)	0.039, 0.171	0.002 (3.1)
Shallow water flatfish	0.02 (0.039)	−0.061, 0.091	0.70 (0.39)

Relative discard varied with the species proportion of the discards; however, vessel, haul size, amount of discard, and species type were not significant factors, nor did they improve model fit using a likelihood ratio test. At low species proportion of catch the variability of the relative difference was higher than at higher species proportion (Fig. 6). This may be an artifact of the greater amount of data available at higher species percentages.

The above analysis is based solely on the point estimates of discard and does not take into account the estimated variance associated with those discard estimates. There were 97 species-specific haul estimates of discard for which the variance of the relative discard estimates was available. These variances were used to scale the relative discards so that each estimate is a normal variate with a mean of zero and variance of one, $N(0, 1)$. Standardized relative discards differed slightly in the shallow water flatfish and arrowtooth flounder fisheries; only two relative discard estimates had associated variance estimates in the Rockfish Program fishery (Fig. 7). Under the null hypothesis that the relative discard difference is equal to zero, 5% of the observations are expected to be outside of the 2.5 and 97.5 percentiles of the standard normal distribution. This larger number of relative discard difference estimates outside these percentiles provides evidence against the null hypothesis. The tails of the distribution are longer than expected with 33 observations (34%) outside the normal upper and lower 2.5% critical values (Fig. 8). These observations were from the arrowtooth flounder fishery (22 estimates, 38% of arrowtooth observations), the shallow water flatfish fishery (10 estimates, 37% of shallow water flatfish estimates), and Rockfish Program (1 estimate, 50% of Rockfish Program estimates). This analysis incorporates the variance associated with the point estimates and supports the conclusion of the previous analysis that there are differences in the estimates between the two methods.

3.6. Comparison of variance estimates

There were 761 paired species-specific haul estimates generated from data collected during this study, one set of estimates for each species and haul. For 64% of standard method discard estimates (488) and 41% of alternate method discard estimates (312) variances could not be computed.

There were 111 species-specific haul estimates of catch for which we had discard variance estimates for both sampling methods. PSE estimates were not consistently larger using one method over the other (Table 5, Fig. 9). Wilcoxon signed rank tests were used to test whether the median of the difference in PSE was equal to zero. The distributions of PSE differences were not centered on zero for the arrowtooth flounder fishery, nor for the dataset as a whole (Table 5). Modeled results were similar with PSE differences

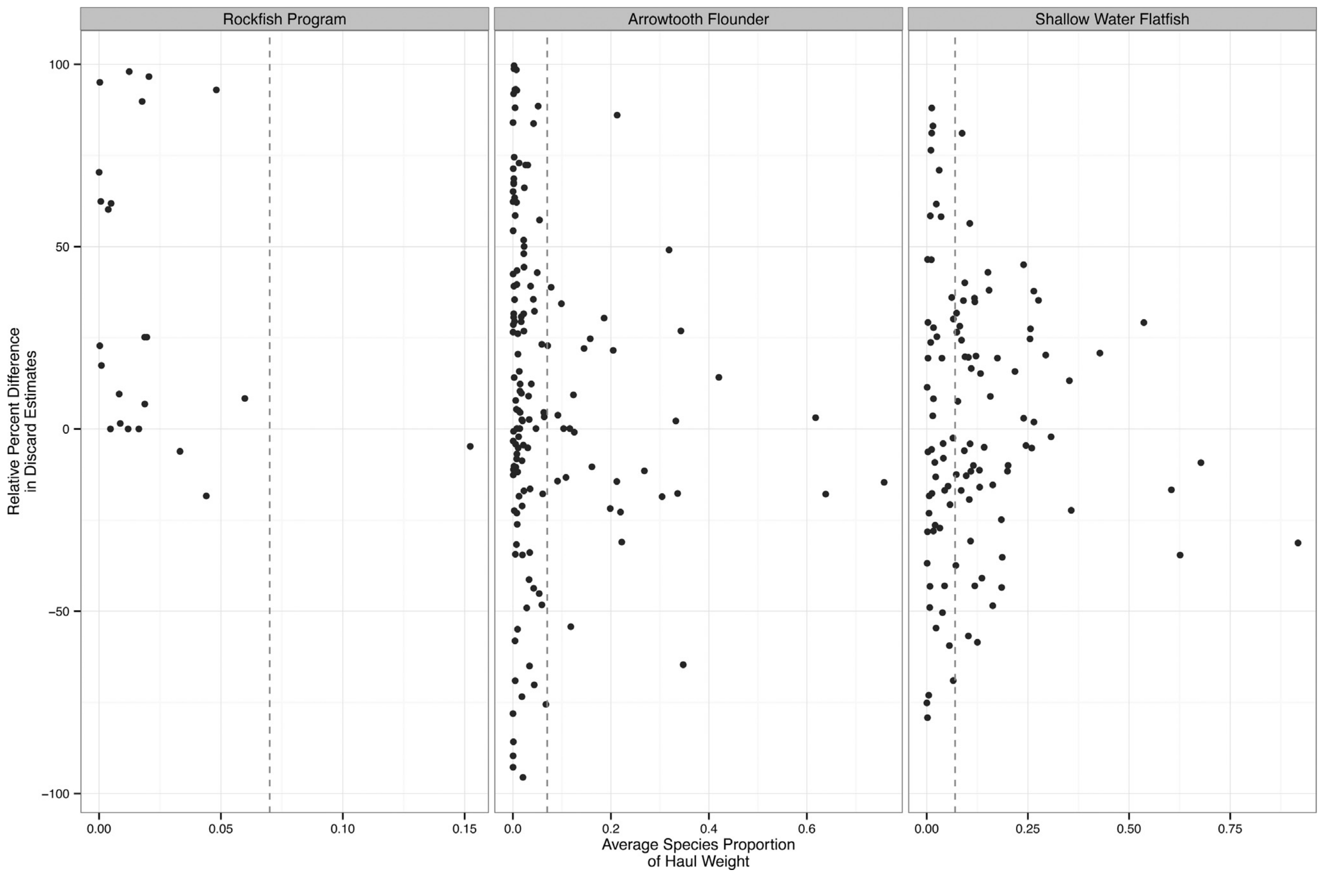

Fig. 6. Relative discard as a function of discard-species proportion of the catch. Dashed line references a species proportion of 7%.

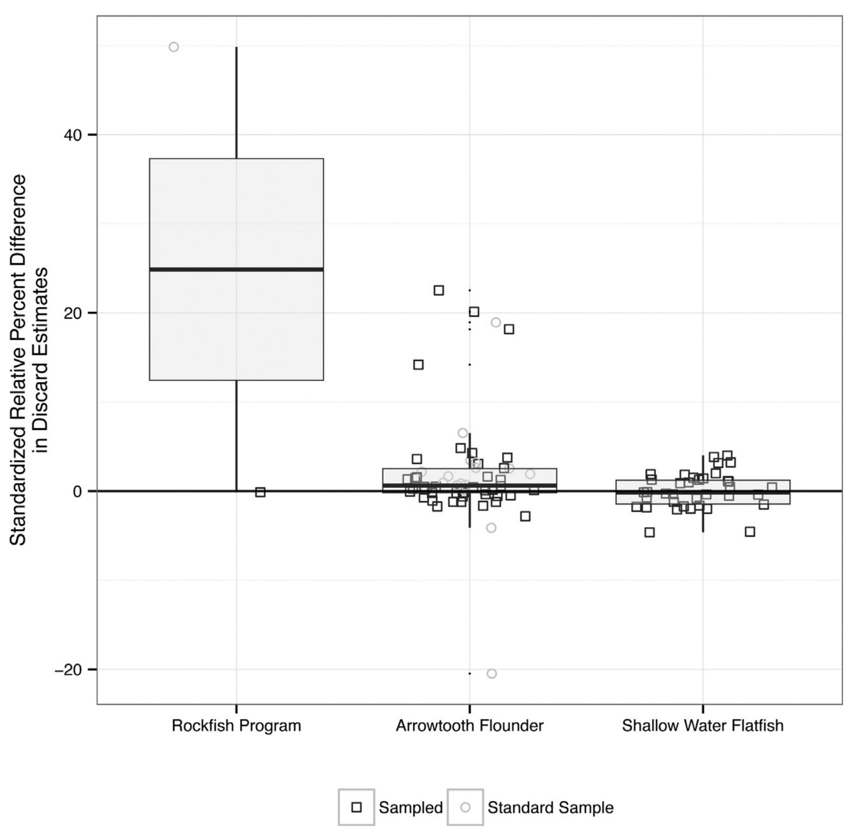

Fig. 7. Standardized relative discard estimates for each fishery. Data points are overlaid onto boxplot. Shape of data points indicates how data were collected: both observer sampled (sampled), standard observer sampled and alternative observer weighed all discards (standard sample), or both observers weighed all discards (weighed).

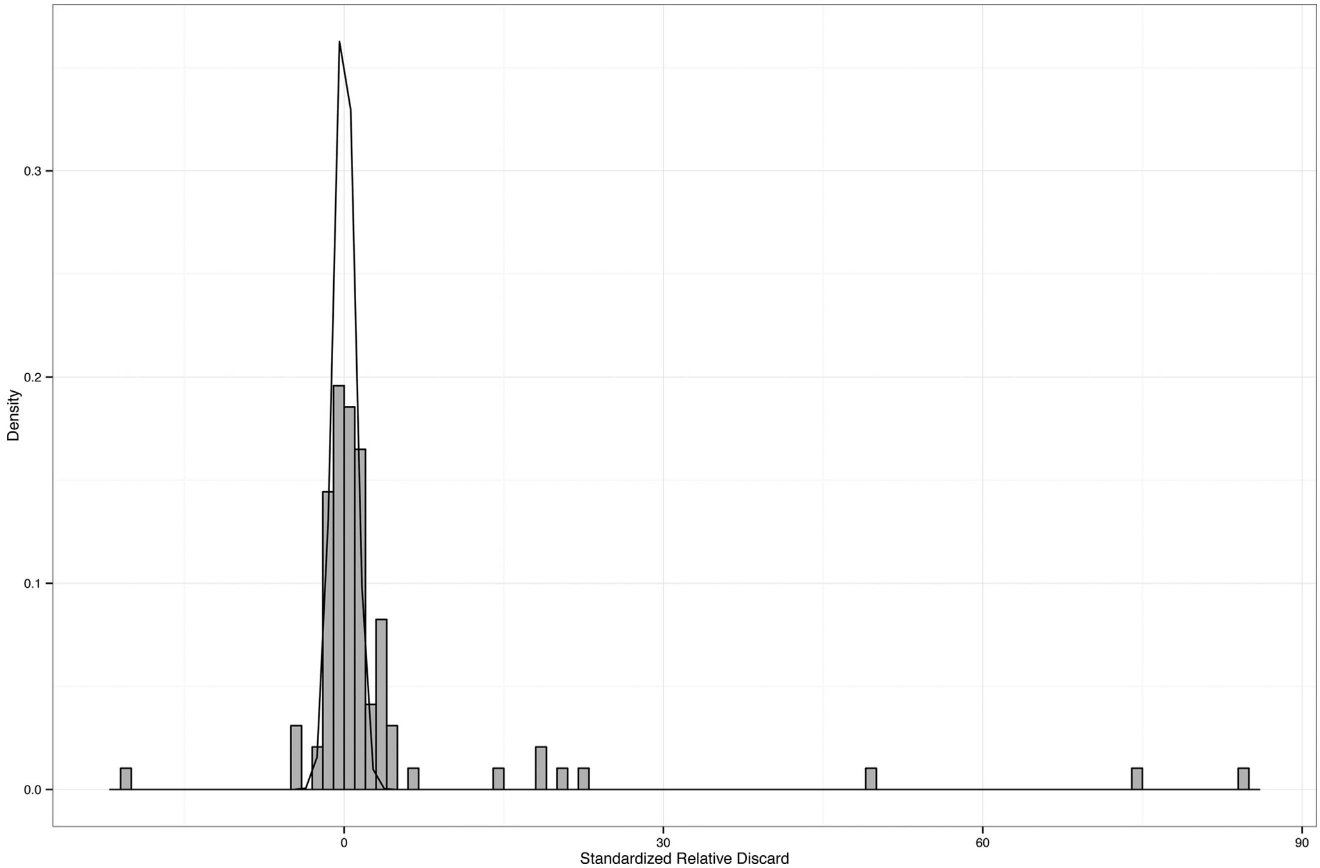

Fig. 8. Distribution of standardized relative discard estimates. Solid line is the standard normal distribution, $N(0,1)$.

Table 5
Mean, median, and range of PSE differences. Negative values indicate that the Alternative method PSE was larger.

Fishery	Rockfish Program	Arrowtooth Flounder	Shallow water flatfish	Total
Mean (Std. Err.)	29.72 (16.27)	28.47 (4.8)	6.54 (6.56)	21.23
Median	0.00	8.54	0.96	6.40
Range	0.00 to 99.6	−74.4 to 99.16	−67.21 to 98.89	−74.4 to 99.6
Wilcoxon statistic, *p*-value	3, 0.37	1482, «0.001	412, 0.36	3730, «0.001
Glm F-statistic, *p*-value	1.8, 0.07	5.9, «0.001	1.0, 0.32	NA

significantly different from zero only for the arrowtooth flounder fishery while species type, vessel, species proportion, and discard size were not significant and did not increase the overall model fit.

Observations where the PSE difference was negative and hence the alternative method PSE was larger tended to be where both the standard method observer and alternative method observer sampled; i.e., neither method resulted all discards being weighed for that haul (Fig. 9, square data points).

The alternative method PSE was lower on more hauls than the standard method PSE in the arrowtooth fishery where the alternative method sample fractions tended to be higher (in some cases the observer was able to weigh all the discards). However, the number of hauls with lower PSE was more evenly distributed between the two methods in the shallow water flatfish fishery, where neither observer was able to completely weigh the discards (Table 5, Fig. 9).

4. Discussion

An increased focus of fishery management on individual vessel accountability has increased the need for higher precision of weight estimates for less common and fishery-limiting species that are discarded at sea. While the common solution is to mandate full (100%) observer coverage to reduce uncertainty in estimation at the trip level (Branch et al., 2006; Fina, 2011), variance that derives from on-deck methods (e.g. within haul) can be substantial and also needs to be properly addressed (Borges et al., 2004). Commercial fishing vessels are not perfect sampling platforms and all sampling methods suffer from implementation issues that potentially impact the quality of the data collected. Sampling methods that can be easily implemented in the commercial fishing environment while minimizing departures from the sample design will result in unbiased data collections and estimation. This study shows that sampling methods differ in their ease of implementation and that estimates based on data collected using these methods vary in their precision estimates.

Evaluations of catch estimation methods, sampling methods, effective use of observers' time and effort, and examinations of sampling bias have been conducted in other fisheries. For example, comparisons between observer data and logbook data have been used to investigate the nature of discrepancies in total catch, areas fished, and trip duration within the fisheries of the northeast United States (Rago et al., 2005), New Zealand (Bremner et al., 2009), and the North Sea (Borges et al., 2008). These studies did not, however,

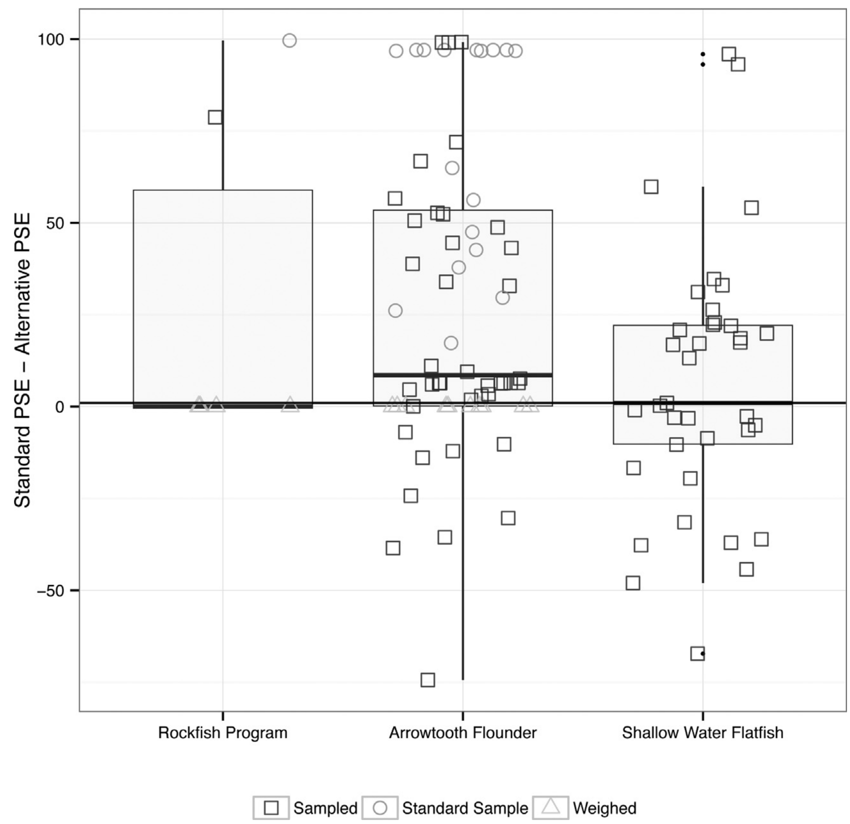

Fig. 9. Distribution of PSE differences for each fishery. Negative values indicate that the Alternative method PSE was larger. Shape of data points indicates how data were collected: both observer sampled (sampled), standard observer sampled and alternative observer weighed all discards (standard sample), or both observers weighed all discards (weighed).

evaluate haul-specific sampling methods or within-haul variability of the catch estimates.

Tamsett et al. (1999b) found that a haul-level sampling method that worked well in one area did not perform well in another area because of crew sorting and discarding behavior precluded sampling of all discards. Evaluation of haul-level sampling and discard estimates showed that there was within-haul variability in species catches. In a subsequent study (Tamsett et al., 1999a), a simultaneous comparison of two sampling methods was conducted and no evidence of bias in the estimates was found (no difference in estimates from the two methods), leading to a proposal that the sampling method most appropriate to vessel operations be used. In Alaska, Conners et al. (2009) examined the effects of observer sample fraction under the standard sampling method on resultant variance and potential bias in total catch estimates for individual species on Bering Sea catcher processors. They found decreased estimated variance with increasing sample fraction and no evidence of sampling bias. Few other studies have evaluated within-haul sampling methods or estimation routines for trawl vessels.

Both standard and alternative sampling methods suffer from implementation difficulties. The efficacy of observer sampling methods was affected by the magnitude and the species diversity of discards. Observers were able to contain, sort, and weigh all the catch that was to be discarded in cases where the discard volume was minimal. Impacts to the fishing operations were minor and observers using the alternative sampling method had ample time and space to conduct their work. At the other extreme, the observers were not able to track, isolate, or sample all the discards using the alternative method in high volume discard and high species diversity fisheries. This may be one cause of the differences found in discard estimates between the two methods.

Logistical constraints to sampling are not always under the control of the observer program. Generally, observers using standard methods on trawl catcher vessels are not able to explicitly define sample units and randomly sample from these, and it is difficult to obtain multiple samples from each haul. In this study, observers were not consistently able to collect multiple samples, resulting in an inability to generate estimates of uncertainty for the discard estimates. Notwithstanding, in this study observers were able to obtain three samples from each haul on a diverse and high volume discard trip; showing that it is possible to impart randomization into their sampling design. On this same trip, the alternative method observer was not able to obtain multiple samples from many of the strata sampled and some strata were not able to be sampled. In addition, while observer samples were often exposed to wave action and observers found themselves sampling in unsafe conditions, this occurred more often for observers using alternative sampling methods. While sampling methods similar to our alternative method have been successfully implemented elsewhere in the United States (NWFSC, 2015), in Alaska, these logistical constraints obstructed sampling activities in some situations and as a result may have introduced sampling biases to the resulting discard estimates. Integrating the new methodology with fishing operations by limiting the amount of at-sea discard or changing the on-deck sorting practices would be necessary to mitigate this potential bias in Alaska.

In this study, observers using the alternative method did not collect the biological data (e.g., length measurements, otoliths) that is part of the standard methodology. On low discard volume and low diversity hauls, collection of this information in the alternative sampling method would have been possible. However, changes to the alternative method such that these data are collected will need to be developed if this method were to be used in Alaska fisheries.

The alternative method does not provide data from which estimates of retained catch can be generated independently from industry reports, hence managers rely on landings reports which do not include haul-specific information such as species composition of retained catch at specific locations or total haul size. Spatial analyses of catch must then rely on assumptions about how much and where retained catches are taken. This loss of spatial resolution can make analyses of fisheries management areas or allocation of catch to specific areas more challenging.

For vessels delivering catch to shoreside processors in Alaska, such as those that participated in this study, the retained (landed) catch is sorted to species or species groups at the shoreside processing facility and catch weights are recorded in landing reports (fish tickets). Although in Alaska these data are not independently verified (as they are in some West Coast fisheries) and landing report accuracy could be improved by shoreside observer monitoring (Faunce et al., 2015), these data can provide reliable estimates of retained catches for most species and species groups at the spatial resolution of the trip.

Different patterns of discard were found in each of the fisheries when detected by both sampling methods. In the Rockfish Program fishery, the discard estimates from the standard method were consistently larger than those generated under the alternative method. This particular comparison is interesting since the observer using the alternative method weighed all the discards in this fishery and the observer using the standard method sampled from the total catch. Compare this to the situation in the shallow water flatfish fishery where both the observers sampled from either the discards or the catch and the standard method discard estimates were on average smaller than the alternative method estimates. In this case, the relative differences ranged from plus to minus 100%, and the pattern was much less pronounced. These differences may be due to imperfect implementation of the alternative method, assumptions made during the estimation process, or biases in the data collection.

The PSE under the alternative method was smaller than that for the standard method when either the observer using the alternative method was able to completely weigh all the discards or in some cases where both observers were sampling (not able to weight all discards). The higher PSE estimates observed for the alternative sampling method estimates may be partially explained by the underestimation of the standard method variances. Both the total haul weight and the percent of the species-specific catch that is discarded are estimates, and thus subject to some variability; however this variance is not quantified and hence not included in the overall PSE estimate. In situations where the alternative method was more difficult to implement, neither sampling method had consistently smaller PSEs, suggesting that the alternative sampling method does not always produce data that generates lower variance discard estimates.

The alternative method addresses problems that are inherent in the standard observer sampling methods in Alaska. For example, when the species being discarded is a small portion of the overall catch, the generally small sample fraction that is inherent in the standard sampling methods may lead to detection problems and high variance discard estimates. Compounding this problem is the estimation of the percent of species catch that is retained by the vessel. This percent-retained estimate is used to estimate the amount of discard for a particular species. However, it is not based on sampling but rather on the observer's observations and knowledge of the fishing operations. For many species, the vessel either retains or discards the species catch entirely (percent retained is zero or 100%). However, for some species the portion of the catch that is damaged, unmarketable, or above the maximum retention allowance is discarded. In this latter case, the visual estimate of the percent retained is used to estimate the retained and discarded portion of the total catch. In contrast, the alternative sampling method does not rely on this visual estimate of percent retained. The detection of species at the lowest proportions was greater using the alternative method. The percent retained, used to estimate the discard in the standard method, is not required since discards are sampled directly. Additionally, the measure of the total population size (haul size or total discard amount) differs between the two methods. Standard methods entail the observer measuring the volume of the catch, estimating the density of the catch, and using these to determine the total haul size. Under the alternative sampling method, the total amount of discard is estimated directly by weighting all discards or estimating a volume of discard and defining sample units in terms of volume (no need to estimate the density of catch). Lastly, the alternative method sample fraction is based solely on the at-sea discard amount, hence the same total sample weight represents a larger sample fraction and detectability problems are reduced. In cases where quotas for rare and uncommon species of bycatch are used in management of the fishery, the ability to detect those discards is critical to quota management.

5. Conclusions

Estimating at-sea discards is an issue that has received considerable attention worldwide. Global and local fisheries catch including discards has been estimated (e.g., Alverson et al., 1996; Rochet et al., 2002; Kelleher, 2005), optimal sample sizes have been determined for programs that employ hierarchical designs (e.g., Allen et al., 2002), factors that are associated with discards have been identified (Feekings et al., 2012), and the best way to estimate catch and discards have been explored (Tamsett and Janacek, 1999; Borges et al., 2004). Despite these important endeavors, we are aware of few studies that compare the perhaps more limited, but no less important relative performance of different on-deck sampling methods under varying catch conditions and fisheries (e.g., Tamsett et al., 1999a,b).

The two sampling methods examined here are fundamentally different in that they are collecting different data, focused on different objectives, and subject to different assumptions. Ultimately the alternative method can be implemented in Alaska without large impacts to fishing operations and without violation of assumptions in some low discard situations. However, given the current fisheries in the Gulf of Alaska, this potential is limited to situations where the at-sea discard of catch is minimal. If fisheries were constrained so that low discard volumes were more prevalent, the alternate method has the potential to provide higher quality discard estimates. In the absence of these management measures, additional work also needs to be conducted on the sample methods before recommendations for a preferred method can be made. For example, the potential biases associated with estimation of total haul size using the standard method should be assessed with the intent of developing more accurate haul size estimates. Investigation of methods that will result in an increase in the total size of sample collected by observers and the number of sampled collected should also be considered.

Although the alternative method appeared to be feasible in some Alaska fisheries, this was not found to be the case in all fisheries tested. Similar to the results of Tamsett et al. (1999a), this highlights the need for comparative field testing of sampling methods under realistic fishing operations, preferably against a known standard such as weighed catch of each species. While our focus is limited to a single level of a hierarchical sampling program, it illustrates the difficulty in applying a "one-size-fits-all" approach to designing observer sampling programs and the need to carefully examine the assumptions behind each level of the sampling hierarchy (see Rochet and Trenkel, 2005).

Acknowledgements

This work would not have been possible without the dedicated professional fisheries observers who served as the project's at-sea scientists or without the voluntary participation of the owners, operators, and crew of the fishing vessels that allowed us onboard. Specifically we would like to thank observers: Robert Davis, Laura Galati, Kenny Hardin, Carla Hilts, Mike Levine, Lauren Loe, Tom Mauer, Dave Pearmain, and Rory Ricks; the vessels *F/V* Caravelle, *F/V* Laura, *F/V* Marathon, *F/V* Mar Del Norte, *F/V* Michelle Renee, *F/V* Pacific Star, *F/V* Sea Mac, and *F/V* Topaz.

Katy McGauley (Alaska Groundfish Data Bank) provided invaluable at-sea support and technical expertise to the project. The programmatic support on which this project depended was provided by the training, data processing, and contract management staff of the Fisheries Monitoring and Analysis Division of the Alaska Fisheries Science Center, in particular Brian Mason, Mike Vector, Bob Maier and Heather Weikart. Saltwater, Inc., provided observers while under contract to NOAA/NMFS.

We are grateful to Jason Gasper, Elizabeth Chilton, and Michael Vechter for valuable comments. This analysis also benefitted from discussions with Ben Fissel.

Funding for this project was provided by the North Pacific Research Board (Project 1017, NPRB publication number 566), NOAA/NMFS Alaska Observer Program, and the Pacific States Marine Fisheries Commission.

The findings and conclusions in this paper are those of the authors and do not necessarily represent the views of the National Marine Fisheries Service, NOAA or the Pacific States Marine Fisheries Commission.

References

Alaska Fisheries Science Center (AFSC). 2011. Observer sampling manual. Fisheries Monitoring and Analysis Division, North Pacific Groundfish Observer Program. AFSC, 7600 Sand Point Way N.E., Seattle.
Allen, M., Kilpatrick, D., Armstrong, M., Briggs, R., Course, G., Perez, N., 2002. Multistage cluster sampling design and optimal sample sizes for estimation of fish discards from commercial trawlers. Fish. Res. 55, 11–24.
Alverson, D.L., Murawski, S.A., Pope, J.G., 1996. A global assessment of fisheries bycatch and discards. FAO Technical paper 339. Rome. Available online at: http://www.fao.org/docrep/003/T4890E/T4890E00.htm.
Borges, L., Keekan, O.A., van Helmond, T.M., Couperus, B., Dickey-Collas, M., 2008. What do pelagic freezer-trawlers discard? ICES J. Mar. Sci. 65, 605–611.
Borges, L., Zuur, A.F., Rogan, E., Officer, R., 2004. Optimum sampling levels in discard sampling programs. Can. J. Fish. Aquat. Sci. 61, 1918–1928.
Branch, T.A., Rutherford, K., Hilborn, R., 2006. Replacing trip limits with individual transferable quotas: implications for discarding. Marine Policy 30, 281–292.
Bremner, G., Johnstone, G., Bateson, T., Clarke, P., 2009. Unreported bycatch in the New Zealand West Coast South Island hoki fishery. Marine Policy 33, 504–512.
Cahalan, J.A. 2010. At-sea monitoring of commercial north Pacific groundfish catches: a range of observer sampling challenges. AFSC Quarterly Report Feature, July-August-September 2010.
Cahalan, J.A., Mondragon, J., Gasper, J., 2014. Catch Sampling and Estimation in the Federal Groundfish Fisheries off Alaska: 2015 Ed. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-286, 41 p.
Conners, M.E., Cahalan, J., Gaichas, S., Karp, W.A., Loomis, T., Watson, J., 2009. Sampling for estimation of catch composition in Bering Sea trawl fisheries. NOAA Technical Memorandum NMFS-AFSC-199.
Faunce, C., Cahalan, J., Bonney, J., Swanson, R., 2015. Can observer sampling validate industry catch reports from trawl fisheries? Fish. Res. 172, 34–43, http://dx.doi.org/10.1016/j.fishres.2015.06.007.
Faunce, C., Cahalan, J., Bonney, J., Swanson, R., 2013. Alternative catch monitoring of Alaskan groundfish. NPRB Project 1017 Final Report. 65 pgs. plus appendices.
Feekings, J., Bartolino, V., Madsen, N., Catchpole, T.L., 2012. Fishery discards: factors affecting their variability within a demersal trawl fishery. PLoS One 7, e36409.
Fina, M., 2011. Evolution of catch share management: lessons from catch share management in the North Pacific. Fisheries 36, 164–177.
Jessen, R.J., 1978. Statistical Survey Techniques. John Wiley and Sons, New York.
Kelleher, K. 2005. Discards in the world's marine fisheries: an update. FAO Fisheries Technical Paper. 470. 131 pp.
McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models, 2nd ed. Chapman & Hall/CRC, 511 pp.
Northwest Fisheries Science Center (NWFSC), 2015. West Coast Groundfish Observer Program 2015 Non-Catch Share Training Manual. West Coast Groundfish Observer Program. NWFSC, 2725 Montlake Blvd. East, Seattle, Washington, pp. 98112.
Development Core Team, R., 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org.
Rago, P.J., Wigley, S.E., Fogarty, M.J., 2005. NEFSC bycatch estimation methodology: allocation, precision, and accuracy. Northeast Fish. Sci. Center Ref. Doc., 05–09.
Rochet, M., Péronnet, I., Trenkel, V.M., 2002. An analysis of discards from the French trawler fleet in the Celtic Sea. ICES J. Mar. Sci. 59, 538–552.
Rochet, M., Trenkel, V.M., 2005. Factors for the variability of discards: assumptions and field evidence. Can. J. Fish. Aquat. Sci. 62, 224–235.
Tamsett, D., Janacek, G., 1999. Sampling trips for measuring discards in commercial fishing based on multilevel modelling of measurements in the North Sea from NE England. Fish. Res. 42, 103–115.
Tamsett, D., Janacek, G., Emberton, M., 1999a. A comparison of methods for onboard sampling of discards in commercial fishing. Fish. Res. 42, 127–135.
Tamsett, D., Janacek, G., Emberton, M., Lart, B., Course, G., 1999b. Onboard sampling for measuring discards in commercial fishing based on multilevel modelling of measurements in the Irish Sea from NW England and N Wales. Fish. Res. 42, 117–126.
Thompson, S.K., 1992. Sampling. John Wiley & Sons, Inc., New York.