ESTIMATING WEIGHT AND IDENTIFYING SPECIES THROUGH ELECTRONIC MONITORING (EM):

A PRELIMINARY COMPARISON OF ELECTRONIC AND OBSERVER-BASED REPORTING

Interim Technical Review
Option Period I
Contract EA133F-10-SE-0949

Prepared for:
U.S. Department of Commerce
National Oceanographic and Atmospheric Administration
National Marine Fisheries Service
Northeast Fisheries Science Center
Woods Hole, MA 02543

06 September 2012

Prepared by:
Maria Jose Pria
Howard McElderry
Andrew Fedoruk
Randy McVeigh

Archipelago Marine Research Ltd.
525 Head Street
Victoria, British Columbia
V9A 5S1, Canada
www.archipelago.ca

TABLE OF CONTENTS

ABSTRACT

Exploratory experiments were carried out between November 2011 and May 2012 as part of the New England electronic monitoring (EM) pilot project to provide preliminary data on: 1) whether estimating weight using length-weight conversions and/or volumetric estimates should be pursued further and 2) whether EM data can be used to identify discarded fish species in the Northeast (NE) groundfish fishery.

Observers and EM systems were simultaneously in place and the experimental methods were intended to allow for comparisons of observer and EM reviewer data at the individual fish and basket level. EM data was analyzed by two independent reviewers, Reviewer A and Reviewer B.

Results from the length-weight experiments found no significant difference between weights calculated from observer lengths and actual weights collected by the observer (bootstrap mean difference of -0.3491 lbs and upper bound of 95% CI 0.2247 lbs). Bootstrapped mean differences between observer and reviewer calculated weights revealed a statistically significant difference between the measurements and, on average, flounder species were underestimated by 3.67% (Reviewer A) and 8.02% (Reviewer B). Atlantic cod was, on average, underestimated by 8.78% (Reviewer A) and overestimated by 12.20% (Reviewer B).

Volume experiment results showed that reviewers overestimated the weight of flounders compared to actual weights on average by 1.778 lbs (Reviewer A) and 0.872 lbs (Reviewer B) per basket using small baskets and by 4.850 lbs (Reviewer A) and 6.32 lbs (Reviewer B) per basket using bushel baskets. Gadids were overestimated on average by 1.79 lbs (Reviewer A) and 2.40lbs (Reviewer B) by basket. Reviewers overestimated mainly due to very low volumes being rounded up to ¼ basket full estimates.

Species identification experiment results showed that Reviewer A and B had similar results for sand dab flounder, Atlantic cod and ocean pout, where reviewers matched observer identification in >90% of entries. American plaice flounder had 63% and 66% matched identification for Reviewer A and B respectively. American plaice flounder was difficult to identify whenever its mouth was not clearly visible. Yellowtail flounder and winter flounder had high matching success for Reviewer A (97% and 91% respectively) but Reviewer B had difficulties detecting the identifying characteristics on these species (66% and 19% matching success respectively). Expansion of the exploratory experiments is needed to include additional species and increase the sample sizes for others as well as to incorporate methodology changes to increase the identification success for American plaice and achieve greater consistency in identifying winter, and yellowtail flounder among reviewers.

The data collected during these exploratory experiments was sparse and the results presented in this report are preliminary. However, the preliminary results show that the use of length-weight relationships should be pursued further as it is a promising method for estimating discarded weight of some regulated species and that the EM video can be used by reviewers to consistently identify a variety of species, including some, but not all, flounders. Volumetric methods may not be well suited for accurate weight estimation

in applications with low catch volumes since reviewers overestimated weight mainly due to very low volumes being rounded up to ¼ basket full estimates. Overall the results are positive, especially considering that there was limited opportunity to improve the methodologies since the experiments spanned only 14 trips, and we do not see any obvious obstacles to working on resolving the outstanding issues identified through further work. This work should include expanding the experiments in order to collect data on more species, improve species identification and weight estimation as well as develop operational methodologies.

1. INTRODUCTION

In April 2010, the Fisheries Sampling Branch (FSB) of the Northeast Fisheries Science Center (NEFSC) of the National Oceanic and Atmospheric Administration (NOAA) contracted with Archipelago Marine Research Ltd. (Archipelago) on a multi-year project to test the applicability of electronic monitoring (EM) technology for collecting catch and effort data aboard vessels, and evaluating the utility of EM in monitoring catch in the Northeast multispecies fishery (also referred to as the NE groundfish fishery).

The first year of the project focused on building local capacity to support current and future EM efforts in the region and gathering an initial comparative data set with observer and EM data (Pria *et al.*, 2011). For this initial data set, observer and EM data were collected independently and used different methods which introduced too many external factors to be able to understand how species identification and weight estimation using EM data may differ from observer identification and weight collection.

To begin answering the question of how these weight estimation methodologies and species identification may compare between observers and EM reviewers, we carried out exploratory experiments that used observer data to ground truth the estimated weights and species identification gathered from EM. Given that weight cannot be determined directly from EM visual data, we chose to explore two methods for estimating weight from video data: using fish length and applying length-weight relationships, and using volumetric estimates and applying density factors.

The exploratory experiments had two independently tested objectives:

- To provide preliminary data to decide whether estimating weight using length-weight conversions and/or volumetric estimates should be pursued further, and
- To provide preliminary data on whether EM data can be used to consistently identify discarded fish species in the NE groundfish fishery.

The exploratory experiments were carried out on a commercial fishing operation so that the experiments would be based on real catch composition and at-sea environmental conditions. However the exploratory experiment design was as independent of vessel layout or gear type as possible and was based on a semi-controlled environment with the intention to maximize the alignment between the two data sources and limit external factors influencing the comparisons. The design was not intended to adhere to operational observer or EM on-board methodologies.

This report presents the work done during the exploratory experiments with the intention of identifying which methodologies are worth pursuing further. In addition, the report identifies some of the key methodology elements that would be required when developing operational applications.

2. Materials and Methods

Vessel Selection
Vessels used for these exploratory experiments were selected from those participating in the New England EM pilot study. Vessels were selected based on the following criteria:

- Good track record for providing high quality EM data (complete EM data collection for the entire fishing trip with EM system powered from port to port).
- Actively targeting and discarding regulated species.
- Vessel captain agreeable to carrying observers and modifying catch-handling practices for the purpose of the experiments.

Data used in this report were from two day-trawlers, herein referred to as Vessel A and Vessel B.

Data Collection Timeframe
Exploratory experiment data collection took place between November 2011 and May 2012 (Table 1).

Table 1 Data collection periods by experiment for each of the two vessels

	Length	Volume	Species Identification
Vessel A	Nov 2011- Feb 2012	Nov 2011 – Mar 2012	Mar 2012
Vessel B	n/a	May 2012	May 2012

Due to the nature of the data collected for length and species identification experiments (catch had to be sorted by species for length experiment and randomly sorted for species identification experiment), length and species identification data could not be collected on the same haul. Volume data were collected on all experimental hauls. Observers concentrated on collecting length and volume data for the first part of the experimental data collection period and species identification and volume data for the second part. The data collection period finished at the end of May.

Species Involved in the Experiment
The experiment concentrated on working with discarded regulated species. These species are prohibited or regulated through trip limit and ACE.

The species regulated through trip limit was Atlantic halibut (*Hippoglossus hippoglossus*) while those prohibited were Atlantic wolffish (*Anarhichas lupus*), ocean pout (*Zoarces americanus*) and sand dab flounder (*Scophthalmus aquosus*). ACE regulated species were:

- Atlantic cod (*Gadus morhua*)
- Haddock (*Melanogrammus aeglefinus*)
- Pollock (*Pollachius virens*)
- Redfish (*Sebastes* spp)
- White hake (*Urophycis tenuis*)
- American plaice flounder (*Hippoglossoides platessoides*)

- Winter flounder (*Pseudopleuronectes americanus*)
- Witch flounder (*Glyptocephalus cynoglossus*)
- Yellowtail flounder (*Limanda ferruginea*)

Overview of Experimental Design

Experimental design was based on collecting and reviewing EM data and comparing them to data collected by an observer. For this comparison to be meaningful, EM and observer data had to be collected in a manner that maximized alignment.

FSB project staff were on board each experimental trip to collect the data for the exploratory experiments. For the purposes of this report FSB staff on board experimental trips will be referred to as "observers." Observers collected standard at-sea monitor (ASM) program data as well as data specifically for comparison with EM reviewer data. Only experiment-specific data were used for comparison to EM reviewer data.

In all experimental hauls the observer and crew put aside all regulated species that were to be discarded. The observer then sorted this catch by species into baskets and took a weight of each basket, either using a Marel scale or a spring scale. Spring scales were used during hauls in which the motion-compensated scale could not be calibrated properly. The observer then collected data depending on the experiment being carried out (described in sections 2.1, 2.2 and 2.3).

Two independent reviewers examined all fishing events: one Archipelago reviewer (Reviewer A) and two FSB staff (grouped as Reviewer B). All reviewers were trained in NE groundfish identification and EM video review. No catch information from the observer data set was available to the reviewers ahead of video review.

The exploratory experiments assumed observer data and actual weights from observers to be accurate (i.e. no observer measurement error was calculated or considered in the data analysis). However, it is possible that errors within observer data introduced differences between the data sets.

EM System Description

The EM systems used to gather data consisted of a control centre, a user interface (monitor and keyboard), a suite of sensors (including GPS receiver, hydraulic pressure transducer, and a drum rotation sensor), and up to four waterproof armored-dome closed circuit television (CCTV) cameras (Figure 1).

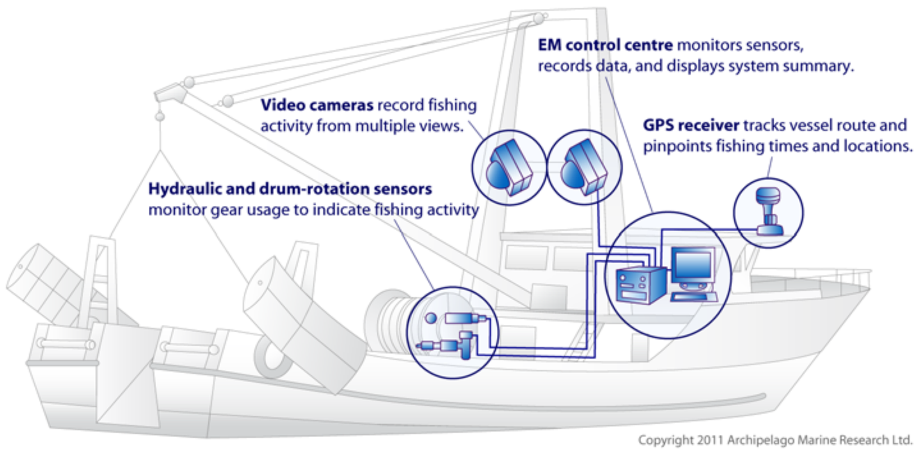

Figure 1 Schematic of the electronic monitoring system used in the experiments

EM Data Collection

EM sensor data were recorded continuously while the EM system was powered, which was intended to be for the entire duration of the fishing trip (i.e. from the time the vessel left port to engage in fishing, to the vessel's return to port). Video recording started once the winch rotated or hydraulic pressure exceeded a threshold level after the vessel left port, and video recording ended when the vessel returned to port.

Camera Configuration

Two cameras were used for recording video data on each vessel (Table 2). One camera was set up to provide an orthogonal, close-up view of the discard area and the other provided a wide-angle view of the deck (Figure 2 and Figure 3). On Vessel A the wide-angle view was initially used for estimating basket fullness (for the first five hauls) but it was determined that a close-up view of the baskets was more appropriate. The wide-angle view was changed at the end of January to capture the observer working area and was used by the reviewer to be alerted when fish entered the close-up camera field of view, which made it easier to align the two data sets at the individual fish-level.

Table 2 Camera installation specifications. Distances are from the camera dome to the center of view. Distance for Vessel A wide-angle view camera corresponds to observer working area view

Vessel	Camera	Lens Size (mm)	Distance (m)	Location
Vessel A				
	Close-up	12.0	2.13	Wheelhouse overhang, starboard side.
	Wide-angle	8.0	3.81	Wheelhouse gantry starboard mast.
Vessel B				
	Close-up	12.0	1.93	Wheelhouse gantry lower crossbar.
	Wide-angle	6.0	3.96	Wheelhouse gantry 'A' frame port side.

Figure 2 Camera views from Vessel A: The top two examples display the initial wide-angle camera view for basket fullness estimation (top left), and the modified wide-angle view for context (top right). The bottom two examples are the close-up view of the discard area used for basket fullness estimation (bottom left) and length and species identification (bottom right)

Figure 3 Camera views from Vessel B: Wide-angle view (left) and close-up view (right)

On Vessel A, the close-up camera view of the discard chute was used for taking length measurements during the length experiments and for identifying catch during species identification experiments. It was also used for estimating basket fullness in ten out of fifteen hauls where volume experiments were carried out. The wide-angle view was used for estimating basket fullness in five out of fifteen hauls.

Similarly, the close-up camera view was used for basket fullness estimation on Vessel B and the wide angle view was used to alert the reviewer to baskets being placed on the close-up field of view.

2.1. Length experiment

All fish in the experiment were presorted by species (to avoid introducing error due to species identification). The fish moved one by one through the close-up view of the discard area, which for the vessel used was a half PVC pipe discard chute, and the discard chute and camera were fixed at a constant distance.

Length measurements were taken according to the NEFSC bottom trawl survey methods (center-line fork lengths for species with forked tails and center-line total lengths for species with round or square tails). The only exception was Atlantic halibut for which NEFSC bottom trawl survey collects forked length, but observers in this experiment collected total length.

Observers measured each fish to the nearest centimetre and placed the fish at the end of the conveyer belt for it to slide onto the discard chute. Fish length measurements were recorded in the same order that the fish were shown on the discard chute to facilitate alignment between the two data sets.

Reviewers determined fish length by measuring the fish in millimeters on the computer screen with a ruler. The screen measurements were then scaled using a multiplier calculated from the reference provided by the graduation marks on the discard chute. The multiplier was calculated by measuring the distance between the marks at the furthest left of the screen, the furthest right of the screen and at the center of the screen and averaging those three measurements. The marks were measured at the middle of the chute.

(Equation 1) Reviewer Length = screen length * multiplier

Graduation marks were adjusted throughout the course of the length experiment data collection period (Figure 4). At the beginning of the data collection period, observers were requested to mark the surface of the chute with regular markings five centimeters apart and on the same plane as that on which the fish were going to be when measured by the reviewers. Observers tried to fulfill this requirement for the first five trips but were unable to due to complications trying to mark the chute while the vessel was at sea. The faint and irregular markings from these trips were used to calculate the scale multiplier but may have affected the reviewer length accuracy for those trips. For the final trip, the captain of the vessel was then requested to mark the chute in a dry location on land, which provided regular 5cm markings.

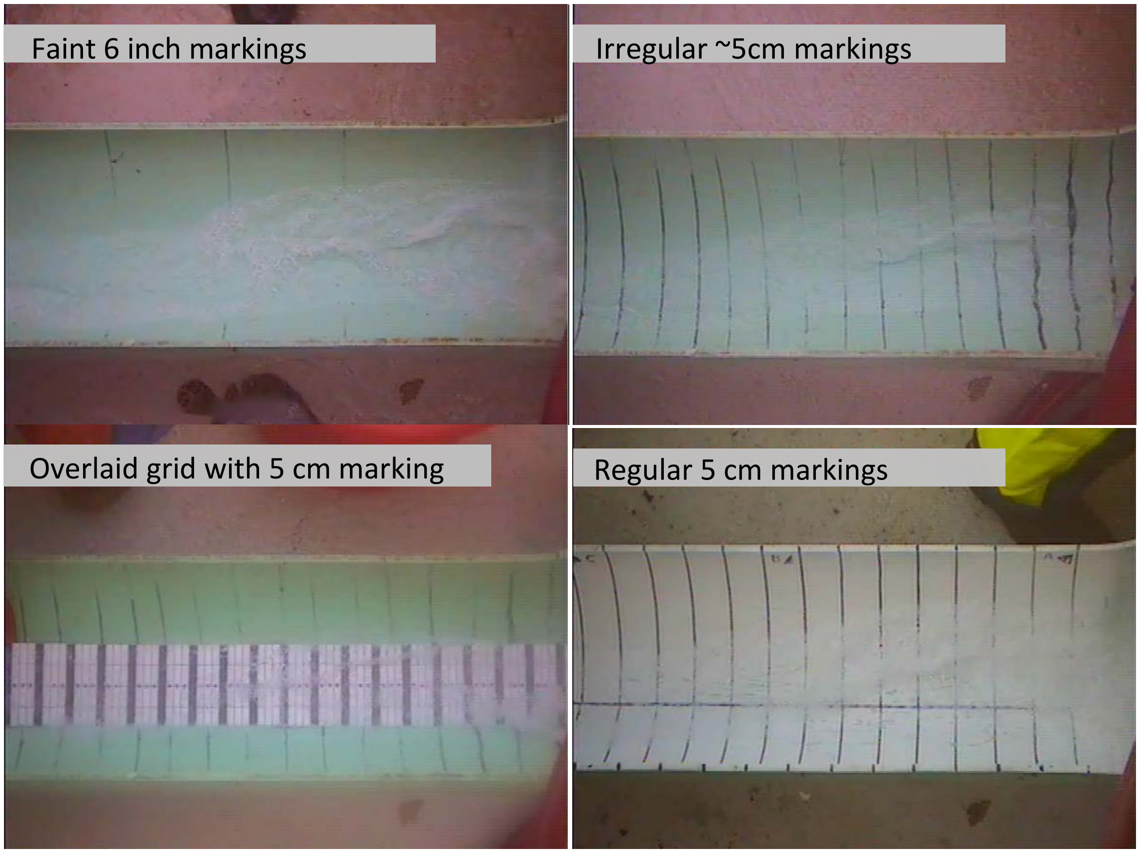

Figure 4 Variation in graduation markings throughout the length experiment data collection

In some instances, reviewers had difficulties measuring the fish length for reasons described below (Figure 5):

- Part of the fish was outside camera view in the images available (referred to as partial image)
- Low image quality caused edge of the fish to be difficult to discern
- Fish curled reducing the two dimensional length on the screen
- Part of the fish covered by the discard chute (chute interference)
- Part of the fish covered by the observer (observer interference)

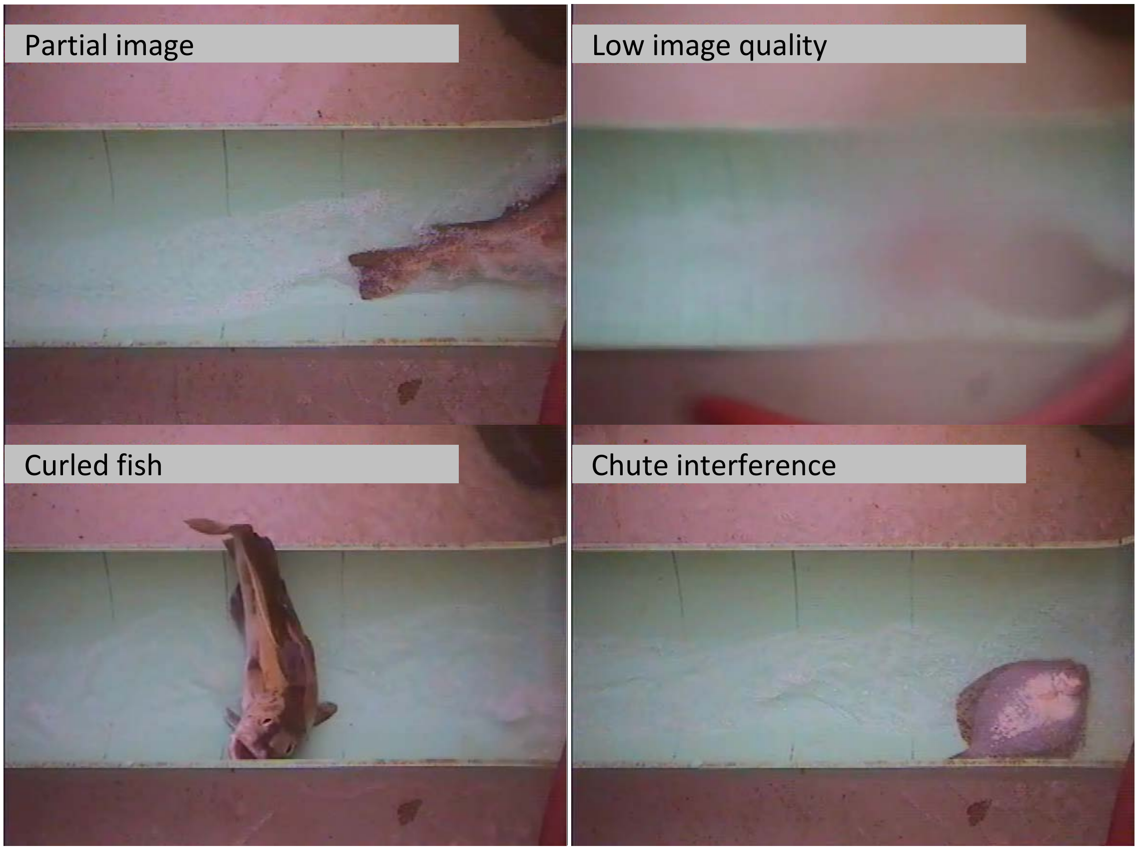

Figure 5 Examples of reasons for identifying a measurement as "compromised" (measuring difficulties)

Observer and reviewer lengths were converted to weights using NEFSC length-weight relationships (Wigley *et al.*, 2003). Length-weight relationships were based on combined sex and on survey data which included winter data for all species, except Atlantic halibut (autumn) and redfish spp (spring/autumn). Fish identified to a species group (i.e. hake, nk or "not known") were not included in the weight comparisons since it was not possible to know which hake species length-weight relationship to apply.

The difference between observer and reviewer lengths and between observer and reviewer weights calculated from length-weight relationships was tested. Furthermore, the validity of the length-weight relationships was investigated by testing whether there was a difference between the observer actual weights and the sum of observer weights calculated from length-weight relationships. Finally, observer actual weights were compared to the total estimated reviewer weight to provide insight into whether this methodology can be applied in an operational setting.

2.2. Volume experiment

All fish were sorted by species into baskets of known volume and all baskets were shown in camera view.

Two types of baskets were used for all but the last exploratory experiment trip: bushel baskets and fish totes. While carrying out these experiments it was apparent that the overall volume of regulated species discards was very low and most of the baskets were not being filled. A third basket type, about half the size of a bushel basket, was sourced at the end of the data collection period and was only used on the last exploratory experiment trip. A description of the baskets and their volumes is provided in Table 3.

Observers and reviewers estimated the fullness of the baskets visually to the nearest ¼ based on how much the fish covered the side of the baskets. Because the bushel baskets had a conical shape and the fish totes had a trapezoidal shape, the volume that corresponded to each height estimate by the reviewer or observer was calculated using geometry.

For all three different types of baskets used, those filled with a small amount of fish were considered to be ¼ full rather than rounded down to 0 and baskets with a fullness height between ¼ and ⅓ were rounded down to ¼, which resulted in the average amount of fish for the "¼ basket estimate" to be 0.1875 rather than 0.25. For this reason the corresponding volume proportion for the "¼ basket estimate" was based on an estimated fullness height of 0.19 for all baskets.

Table 3 Description of the shape and size of the baskets used as well as the corresponding volume for each fullness level estimated by reviewers and observers

Basket Description	Estimated Fullness by Height	Corresponding Volume Proportion	Volume (ft^3)
NEFOP standard conical bushel basket			
	Full	1	1.49
	¾	0.71	1.05
	½	0.44	0.66
	¼	0.15	0.22
NEFOP standard trapezoidal fish tote			
	Full	1	2.69
	¾	0.74	2.00
	½	0.49	1.32
	¼	0.18	0.48
Rectangular small basket			
	Full	1	0.61
	¾	0.75	0.46
	½	0.5	0.31
	¼	0.19	0.12

Reviewer estimated volumes were converted to weights using approximate density factors using Equation 2. It was not possible to obtain independent density or average basket weights for regulated species. Actual weights from full baskets collected in the experiment were used to estimate an approximate density factor. Atlantic cod, haddock and pollock were grouped under "gadids" and were assumed to have the same density for the purpose of these exploratory experiments. Similarly, all flounders were assumed to have the same density. The rest of the species and species groups, including Atlantic halibut, were assumed to have significantly different densities to gadids and flounders and were not included in the volume experiment analysis because there were no full baskets to calculate their approximate density factors.

Equation 2 Weight = Density factor * Volume

The difference between observer and reviewer basket fullness estimates was tested as well as the difference between reviewer estimated weights and the observer actual weights.

2.3. Species identification experiment

The observer mixed all the fish and then randomly took each fish from the basket, recorded the species and placed the fish at the end of the conveyer belt for it to slide onto the discard chute. Fish identification was recorded in the same order as the fish appeared on the discard chute to facilitate alignment between the two data sets. All fish in the experiment moved through the discard chute one by one.

Following NEFOP species identification guidelines, reviewers identified the fish to the lowest taxonomical level possible by using a minimum of two identifying characteristics and were free to use any characteristic they considered appropriate for that species. All reviewers used observer training resources to confirm identification characteristics including (Chase and Galbraith, 2004) as well as their previous experience. In addition Reviewer A used a variety of published resources (Gilbert and Williams, 1993; Douglas *et al*., 1999; Froese and Pauly, 2012). In cases where defining characteristics were not visible, the reviewers recorded the fish under the lowest species group for which identifying characteristics were discernible. Reviewers were asked to write down the characteristics used to identify the catch.

Observer catch entries were paired with each of the reviewer's catch entries to compare identification between the two at the individual catch entry level.

2.4. Data Source Pairing

Since the main goal of the exploratory experiments was to compare reviewer to observer at the catch-item or basket level, it was important to appropriately pair the two data sets. Analysis of individual fish or basket data required a data pairing process since the observer and reviewer data sets sometimes did not match up item-to-item. These mismatches were caused when either the reviewer or the observers did not record a fish or basket that was seen by the other data source. Any records that could not be reconciled between the two data sets were excluded from the analysis.

3. RESULTS

3.1. Length Experiment

3.1.1. Data Inventory

Fish length data were collected throughout seven trips comprised of fifteen hauls in total and included eleven regulated species and one species group (hake, not known). There were 74 actual weights collected by observers throughout the length experiment.

Observers collected 1,462 fish lengths and each reviewer collected 1,463. Individual observations were paired between the observer and reviewer data sets, and pairs where measurements were compromised were excluded from the final sample used in comparisons between observer and reviewer data for calculated weight and fish lengths (this process is summarized in Table 4.).

The pairing of observer to each of the reviewer data sets resulted in a total 1,443 length matching pairs between observer and Reviewer A, and 1,444 length matching pairs between observer and Reviewer B (shown under "Total" in Table 4; the "No Measuring Difficulties" column includes the data pairs for which the reviewers did not highlight any problem measuring the fish length).

Finally, the final sample excluded 166 of Reviewer B fish lengths of species with forks (Atlantic cod, haddock and Atlantic halibut) which were measured as total lengths instead of fork lengths. In addition, the total sample also excluded outliers caused by data entry errors (three from the Reviewer A data set and two from the Reviewer B data set). Comparisons between observer and reviewer fish lengths and calculated weights were based on this final sample of matching pairs which excluded all measurements that had measuring difficulties and those where the reviewer recorded the incorrect length type (total length instead of fork length).

Table 4 Length data pairs available for comparison by species or species group

Species	Original Observer Sample	Observer- Reviewer A Matching Pairs			Observer- Reviewer B Matching Pairs		
		Total	No Measuring Difficulties	Final Sample	Total	No Measuring Difficulties	Final Sample
Yellowtail flounder	588	587	459	459	587	510	510
Sand dab flounder	366	361	288	288	362	326	326
Atlantic cod	352	352	223	221	352	210	47
American plaice flounder	72	65	61	61	65	62	62
Winter flounder	58	57	51	51	57	53	53
Ocean pout	11	11	4	4	11	4	4
Haddock	9	4	4	4	4	4	1
Atlantic halibut	2	2	2	2	2	2	1
Hake, not known	1	1	1	1	1	1	1
Witch flounder	1	1	1	1	1	1	1
White hake	1	1	1	0	1	1	0
Redfish	1	1	0	0	1	0	0
Total	**1462**	**1443**	**1095**	**1092**	**1444**	**1174**	**1006**
Total as percentage of original observer sample		**99%**	**75%**	**75%**	**99%**	**80%**	**69%**

Measurements with no difficulties highlighted comprised 75% and 80% of records for Reviewer A and Reviewer B length measurements respectively. In both reviewer data sets the majority of measuring difficulties were due to only part of the fish being captured on the video as it traveled down the discard chute (Table 5). Reviewer A marked more lengths as difficult to measure compared to Reviewer B, and most of the differences between the two were under the “low image quality” category. “Curled fish” was the second most common measuring difficulty for Reviewer B and the third for Reviewer A but did not represent a large proportion of the total measurements (5% and 4% for Reviewer A and Reviewer B respectively).

Table 5 Number of observations removed from the reviewer data sets due to compromised measurements, out of a total of 1443 and 1444 matching pairs for the Reviewer A and Reviewer B data sets respectively

Measuring Difficulty	Reviewer A		Reviewer B	
	Number of Observations	Percent of Total Pairs	Number of Observations	Percent of Total Pairs
Partial image	183	13%	200	14%
Low image quality	77	5%	4	0%
Curled fish	71	5%	56	4%
Chute interference	9	1%	0	0%
Observer interference	8	1%	10	1%
Total	**348**	**25%**	**270**	**19%**

3.1.2. Comparisons

Haul-Level Comparisons of Observer Calculated Weights and Actual Weights

Actual haul weights taken by the observers were compared with the sum of the calculated weights from observer length measurements to test the applicability of the length-weight relationships for estimating discarded weight. Both a histogram of the differences and a scatter plot show the haul weights by species were similar between the two methods (Figure 6).

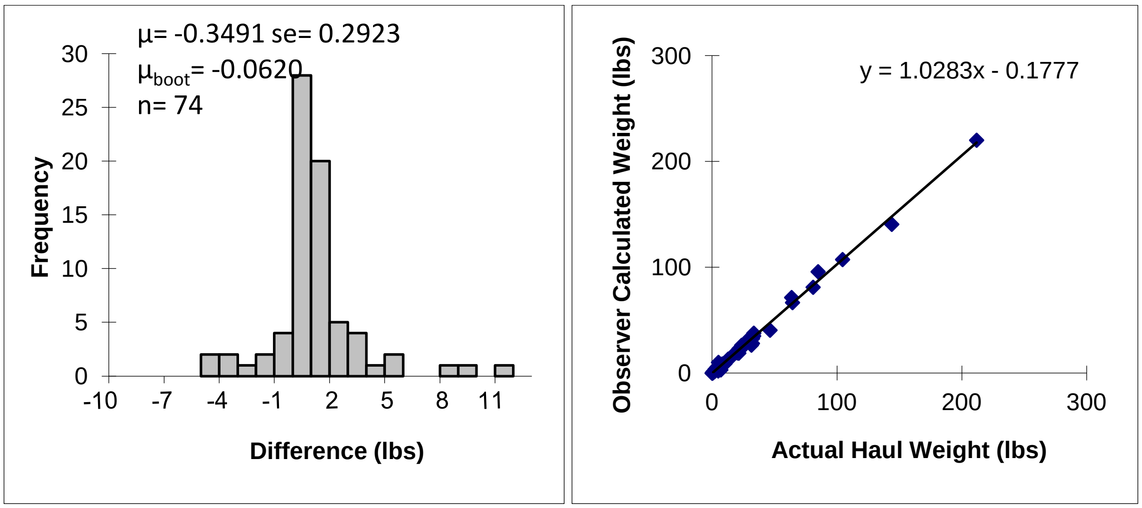

Figure 6 Comparisons of observer-calculated weights and actual weights: the histogram (left) depicts the difference between the observer-calculated weights by species by haul and the actual weights, and the scatter plot (right) illustrates the actual weights and observer calculated weights.

The data were not normally distributed (Shapiro-Wilk p-value<0.01) and could not be transformed using typical functions such as log transform, inverse or cube root. A non-parametric bootstrapping technique and corresponding 95% normal confidence interval could be used in hypothesis testing (Crowley, 1992). In this report it was used for testing for a difference from zero for the means.

There was no evidence of a significant difference on average between observer-calculated weights and the actual deck weights; the bootstrapped haul weight mean difference was -0.3491 lbs with an approximate 95% confidence interval that included zero (-0.9265 lbs, 0.2247 lbs).

Individual Fish Comparisons of Observer and Reviewer Calculated Weights

Having established that the length-weight relationships were an appropriate way of estimating discarded weight by confirming that there was no significant difference between observer calculated and actual weights, comparisons were made between the individual fish weights calculated using length-weight relationships from the observer length data set and each of the reviewer length data sets. Hake, not known was not used for these analyses since a length-weight relationship could not be applied to a species group.

Initial inspection of the data revealed that they were closely correlated (Reviewer A adjusted R squared 0.939; Reviewer B adjusted R squared 0.949) with slopes close to one (Reviewer A slope 0.884; Reviewer B slope 1.135) (Figure 7). Furthermore the distributions of fish weight differences were centered around zero (Figure 7).

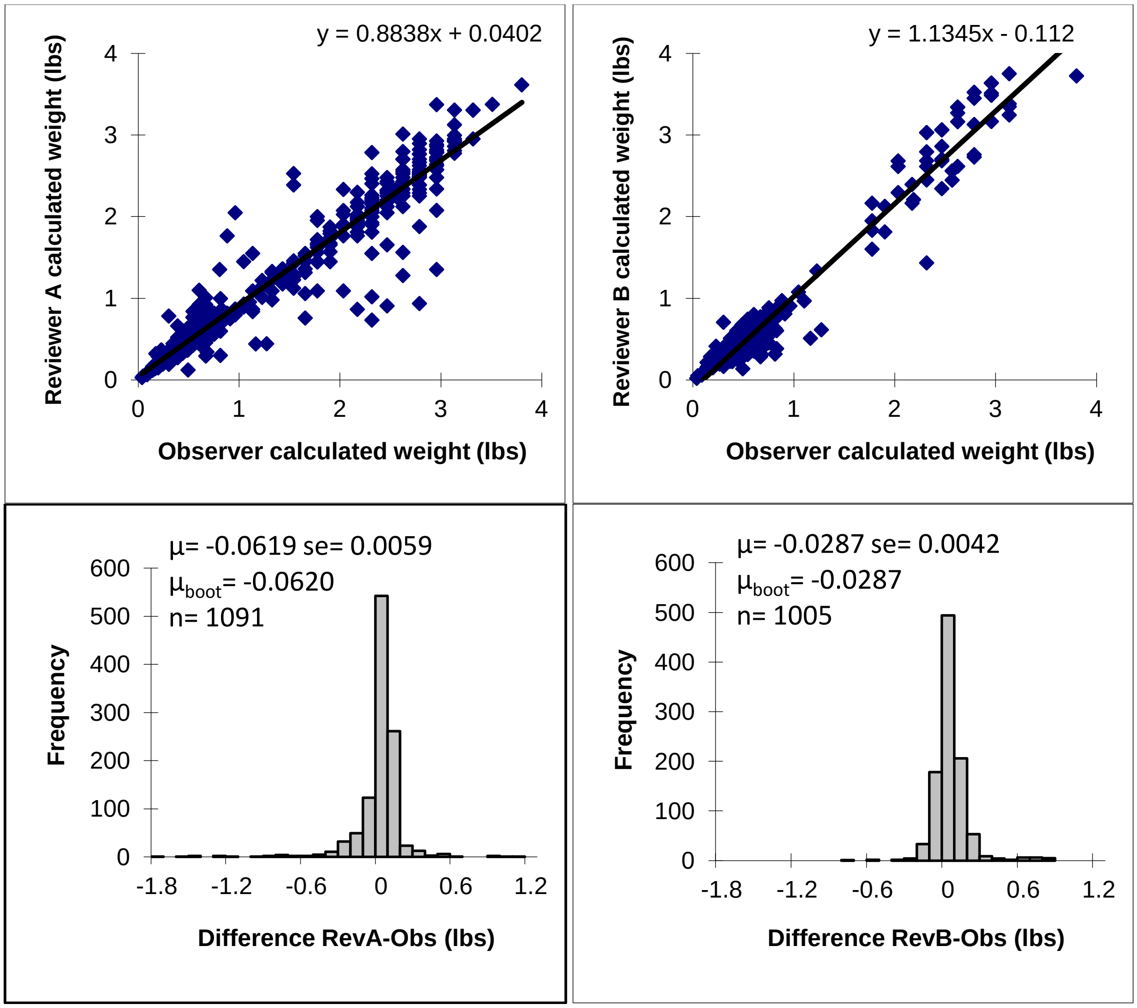

Figure 7 Comparisons of observer and reviewer calculated weights for each reviewer data set (Reviewer A data comparisons displayed on the left and Reviewer B comparisons on the right): The scatter plots (top) show observer and reviewer-calculated fish weights. The histograms (bottom) show the difference between the observer and reviewer-calculated weights.

The differences between reviewer and observer calculated weights were not normally distributed (Shapiro-Wilk p-value<0.01) and could not be transformed. A difference from zero for the means between weights was tested for using the same non-parametric bootstrapping technique as with the haul-level comparisons of observer calculated weight and actual weights. Bootstrapped confidence intervals for Reviewer A and B did not include zero, which provided significant evidence at an alpha level of 0.05 that Reviewers A and B are underestimating fish weight on average (upper bound of confidence intervals were -0.0505 lbs and -0.0203 lbs for Reviewer A and B respectively) (Table 6).

Using only species for which there were more than five records, the results from Kruskal-Wallis one way analysis of variance were significant (Reviewer A H=240.631 and Reviewer B H=94.797, 4 d.f., P<0.001) meaning that the median difference between observer and reviewer calculated weights were different among the five species for both reviewer data sets. This effect was further confirmed using a linear regression model (complete test results included as Appendix I) where there was a major difference between the line of best fit for Atlantic Cod compared to the rest of the species in both Reviewer A and Reviewer B data sets. Furthermore there was no evidence of a major difference among flounder species in both the Reviewer A and Reviewer B data sets suggesting that for all flounder species the measurement bias was approximately the same and these species may be grouped when calculating weight differences between observer and reviewers.

Based on these results, a bootstrapping technique was used to estimate the mean weight difference per fish between observer and reviewer-calculated weights of Atlantic cod and of flounders, which included yellowtail flounder, American plaice flounder, sand dab flounder and winter flounder. Both reviewers underestimated flounder weight as compared to observer calculated weight; Reviewer A by 0.021 lbs, or 3.66% and Reviewer B by 0.045 lbs, or 8.02% per flounder (Table 6).

Atlantic cod was underestimated on average by 0.2172 lbs per fish (or 8.78%) by Reviewer A while it was overestimated by 0.3051 lb per fish (or 12.20%) by Reviewer B.

Table 6 Weight differences per fish calculated from observer and reviewer data sets bootstrapped means and confidence intervals

Species Type	Reviewer A			Reviewer B		
	Mean	Lower 95%	Upper 95%	Mean	Lower 95%	Upper 95%
Overall - lbs	-0.0620	-0.0736	-0.0505	-0.0287	-0.0371	-0.0203
Atlantic cod - lbs	-0.2172	-0.2636	-0.1708	0.3051	0.2130	0.3973
Atlantic cod - %	-8.782	-11.370	-6.193	12.200	8.425	15.980
flounders - lbs	-0.0214	-0.0270	-0.0159	-0.0454	-0.0511	-0.0397
flounders - %	-3.658	-4.656	-2.661	-8.022	-9.044	-7.001

A second Kruskal-Wallis one way analysis of variance, which used fishing trip as the factor to be tested, detected a statistical significant difference between observer and reviewer calculated weight among fishing trips for both Reviewer A and Reviewer B data sets (Reviewer A H=392.299 and Reviewer B H=608.335, 6 d.f., P<0.001).

Individual Fish Comparisons of Observer Lengths and Reviewer Lengths

Comparisons between observer fish lengths and reviewer fish lengths had very similar results as those for the comparisons of calculated weights. Initial inspection of the data revealed that they were closely correlated (Reviewer A adjusted R squared 0.950; Reviewer B adjusted R squared 0.927) with slopes close to one (Reviewer A slope 0.931; Reviewer B slope 1.055) (Figure 8). Furthermore the distributions of fish length differences were centered around zero (Figure 8).

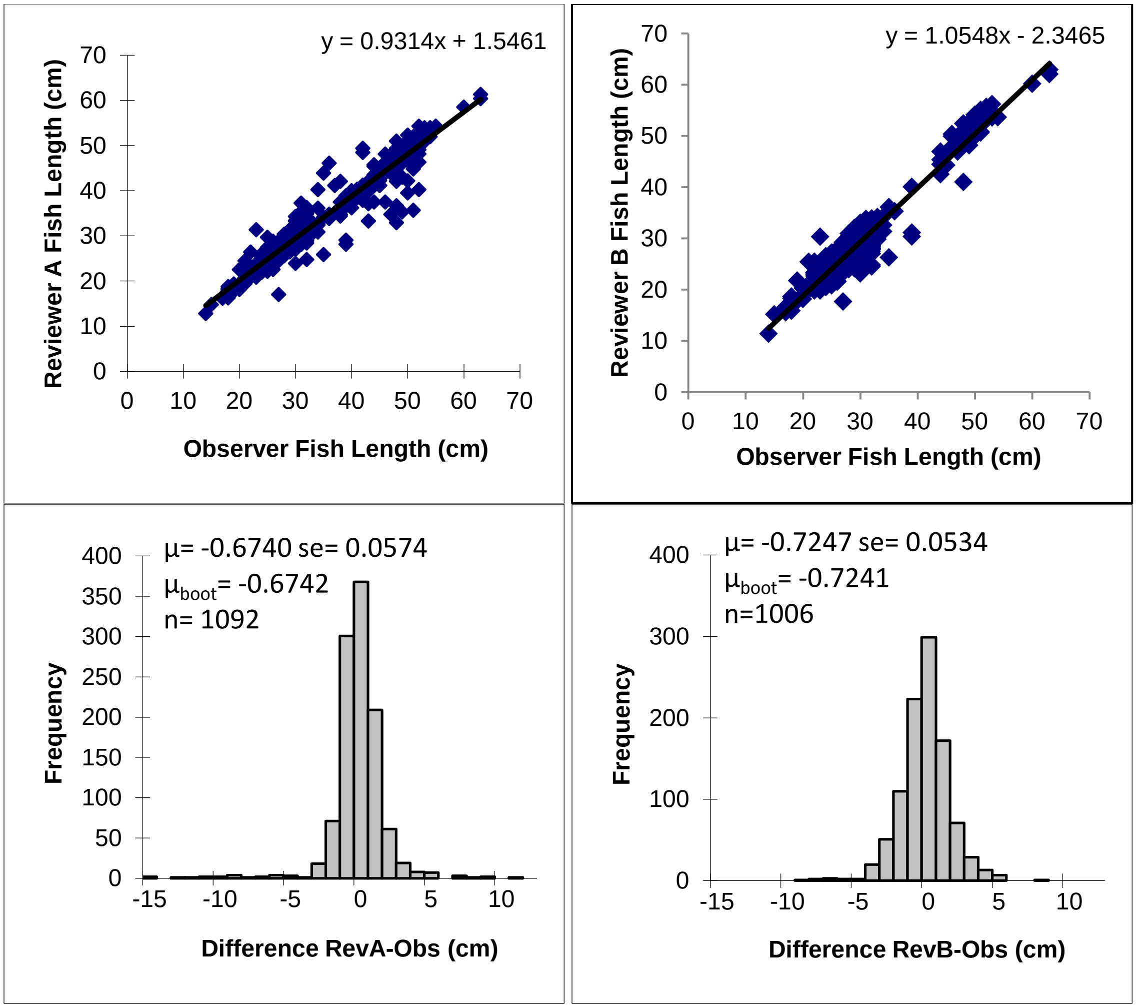

Figure 8 Comparisons of observer and reviewer fish lengths for each reviewer data set (Reviewer A data comparisons shown on the left, and Reviewer B comparisons on the right): The scatter plots (top) show observer and reviewer fish lengths. The histograms (bottom) show the difference between the observer and reviewer fish lengths

The differences between reviewer and observer fish lengths were not normally distributed (Shapiro-Wilks p- value<0.01) and could not be transformed. There was evidence at alpha level 0.05 that the observer and reviewer lengths were statistically significantly different (upper bound of confidence intervals were -0.5620 cm and -0.6183 cm for Reviewer A and B respectively) (Table 7).

Table 7 Difference from measured lengths by observers and reviewers bootstrapped means and confidence intervals

Reviewer	Mean (cm)	Lower 95% (cm)	Upper 95% (cm)
Reviewer A	-0.6742	-0.7865	-0.5620
Reviewer B	-0.7241	-0.8299	-0.6183

Individual Fish Comparisons of Reviewer A and Reviewer B Lengths
Data from the two independent reviews were different. The bootstrapped mean difference between the two independent reviews was -1.3896cm with a 95% confidence interval of (-1.5234cm, -1.2561cm).

3.2. *Volume Experiment*

3.2.1. Approximate Density Factor Calculation

Atlantic cod had 11 baskets estimated as full by reviewers while sand dab flounder and yellowtail flounder each had two. The actual weights of these baskets were used to calculate an estimated basket weight for gadids and flounders and, using the known volume of the baskets, calculate an approximate density for gadids and flounder species groups (Table 8). Since none of these baskets were filled completely, the volume was estimated based on 95% fullness. There were no full baskets for any of the other species and species groups to calculate density and these species were not included in any volume experiment weight comparisons.

Table 8 Approximate basket density for gadids and flounders calculated from average actual weights of full baskets estimates

Species Type	Full Baskets	Mean Basket Weight (lb)	Basket Volume (ft^3)	Approximate Density (lb/ft^3)
Gadids	11	74.92	1.399002	53.55
Flounders	4	79.80	1.399002	57.04

3.2.2. Data Inventory

Volumetric estimate data using bushel baskets were collected throughout 14 trips and 38 hauls and included 11 regulated species, and one species group (hake, not known).

Observers collected 188 volumetric estimates and basket weights using bushel baskets. Reviewer A collected 187 and Reviewer B collected 188 volumetric estimates on bushel baskets. Individual bushel basket observations were paired between the observer and reviewer data sets resulting in a total 183 volumetric estimate pairs between observer and Reviewer A and 185 volumetric estimate pairs between observer and Reviewer B (shown under total matching pairs in Table 9).

Volumetric estimates using fish totes were collected on one trip throughout two hauls for two regulated species. The observer and Reviewer A collected two volumetric estimates and Reviewer B collected one, which resulted in two volumetric estimate pairs between observer and Reviewer A and one volumetric estimate pairs between observer and Reviewer B (shown under total matching pairs in Table 9).

Volumetric estimate data using small rectangular baskets were collected throughout one trip and six hauls and included four regulated species, including redfish. The observer collected 14 volumetric estimates and basket weights. Reviewers also collected 14 volume estimates. Individual basket observations were paired between the observer and reviewer data sets resulting in a total of 14 volumetric estimate pairs between observer and each reviewer (shown under total matching pairs in Table 9).

Weight comparisons were only conducted on gadids and flounder volume estimates excluding full baskets because these were used to approximate a volume estimate.

Table 9 Volume estimates data pairs available for comparison by basket type and species or species group

Basket Type	Species / Species Groups	Fish Type	Observer- Reviewer A Matching Pairs		Observer- Reviewer B Matching Pairs	
			Total	Weight Comparison	Total	Weight Comparison
Bushel Basket						
	Atlantic cod	gadids	37	26	38	27
	Haddock	gadids	3	3	3	3
	Sand dab flounder	flounders	33	31	33	30
	Yellowtail flounder	flounders	31	29	32	29
	Winter flounder	flounders	31	31	31	31
	American plaice flounder	flounders	25	25	25	25
	Witch flounder	flounders	3	3	3	3
	Ocean pout	other	15	n/a	15	n/a
	Hake, not known	other	2	n/a	2	n/a
	White hake	other	1	n/a	1	n/a
	Atlantic halibut	other	1	n/a	1	n/a
	Redfish	other	1	n/a	1	n/a
	Total for bushel basket		**183**	**148**	**185**	**148**
Fish Tote						
	Atlantic cod	gadids	1	n/a	1	n/a
	Atlantic halibut	other	1	n/a	0	n/a
	Total for fish tote		**2**	**n/a**	**1**	**n/a**
Small Basket						
	Witch flounder	flounders	6	6	6	6
	American plaice flounder	flounders	6	6	6	6
	Redfish	other	1	n/a	1	n/a
	Atlantic halibut	other	1	n/a	1	n/a
	Total for small basket		**14**	**12**	**14**	**12**

Discard volumes by species throughout the experiment were very low, resulting in ~78% of baskets being estimated as ¼ full by observer and reviewers (Figure 9). It was not possible to quantify how many entries were rounded up with the data collected but by using data collected for the other experiments the median number of fish in each basket estimated as ¼ full was four, which indicated that over half of the ¼ full baskets were rounded up.

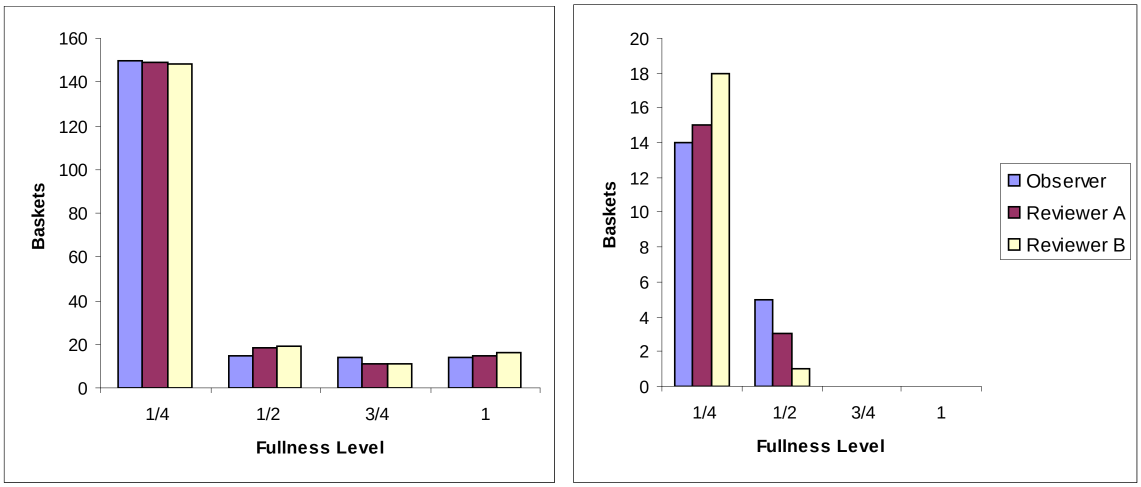

Figure 9 Number of baskets by fullness level from observer and reviewer data sets for Vessel A (left) and Vessel B (right)

3.2.3. Comparisons

Basket Fullness Comparison of Observer and Reviewers Estimates

Bushel baskets fullness estimates between observer and reviewers were identical for 93% and 91% of paired volumes for Reviewer A and Reviewer B respectively (Table 10). Small basket volumes were identical for 71% of paired volume estimates for both reviewers.

Table 10 Frequency of differences between observer and reviewer volume estimates by basket type for each reviewer

	Reviewer A		Reviewer B	
Reviewer - Observer Fullness Difference	Bushel Basket	Small Basket	Bushel Basket	Small Basket
-0.25	4%	21%	3%	29%
0	92%	71%	90%	71%
0.25	3%	8%	5%	
0.5			1%	
0.75	1%		1%	
Total	**183**	**14**	**185**	**14**

Basket Weight Comparison of Actual Weights and Reviewer-Estimated Weights

Both reviewers on average overestimated weight as compared to actual weights for gadids and flounders in both vessels (Figure 10). Both reviewers overestimated weight compared to actual weights for gadids and flounders in both container types. Mean weight differences were greatest for flounders using bushel baskets (5.469 lbs per basket for Reviewer A and 6.323 lbs per basket for Reviewer B). The mean difference was much smaller when the small rectangular baskets were used (1.778 lbs per basket for Reviewer A and 0.872 lbs difference per basket for Reviewer B).

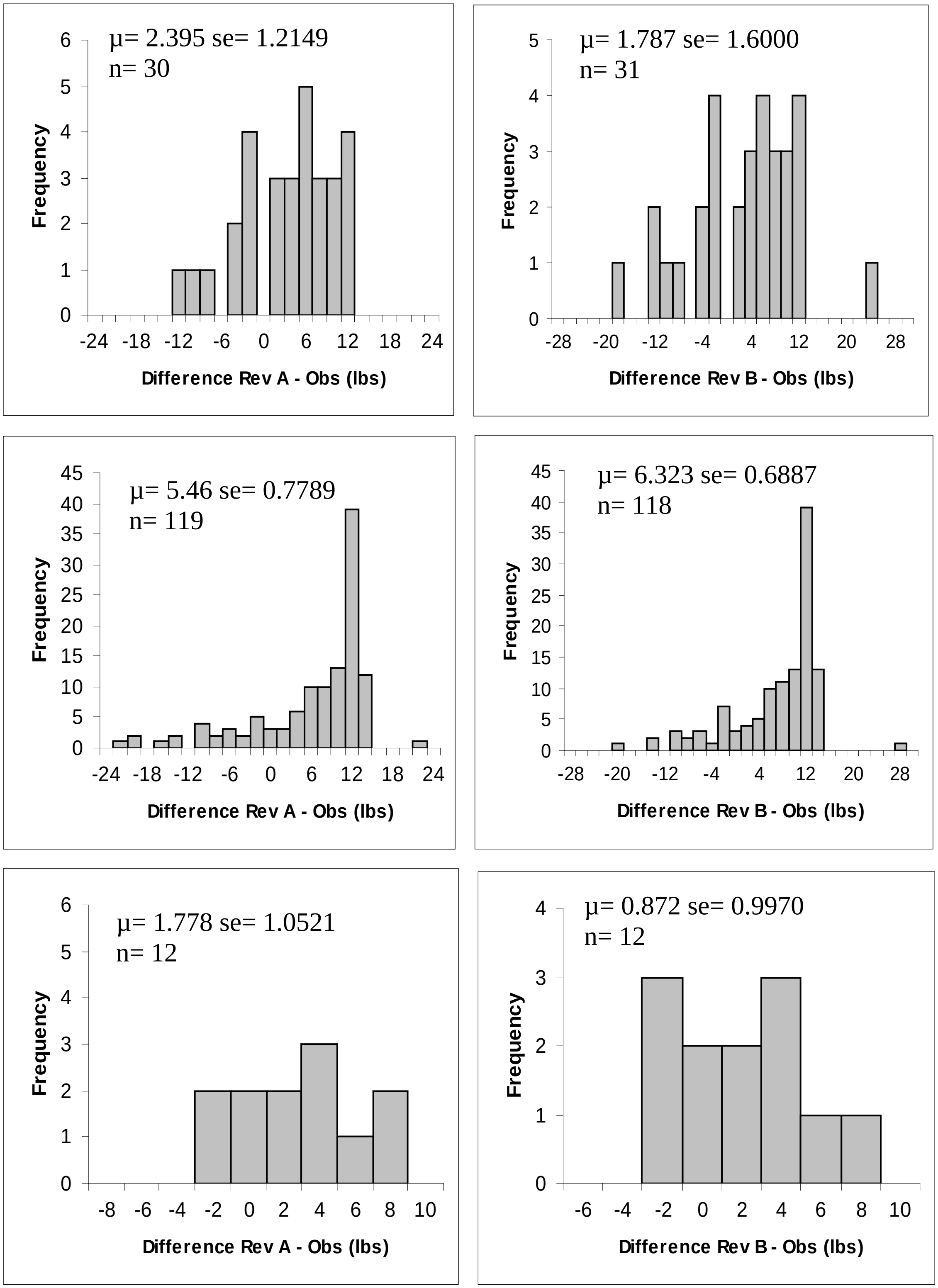

Figure 10 Histograms of the difference between Reviewer A (left) and Reviewer B (right) estimated weight and actual weights (lbs). Bushel baskets gadids (top), bushel baskets flounders (center), small baskets flounders (bottom)

Basket Fullness Comparison of Reviewer A and Reviewer B Estimates
Comparison results of basket fullness estimation between reviewers show a very similar distribution as observer and reviewer comparisons with most paired estimates being identical for both reviewers: 95% of the bushel basket paired entries and 86% of the small basket paired entries (Table 11).

Table 11 Frequency of differences between reviewers basket fullness estimates by vessel.

Reviewer A – Reviewer B Fullness Difference	Bushel Basket	Small Basket
-0.75	1%	
-0.25	3%	
0	95%	86%
0.25	1%	14%
0.5		
0.75		
Total	**185**	**14**

3.3. Species Identification Experiment

3.3.1. Data Inventory

Species identification data were collected on Vessel A throughout six trips consisting of 19 hauls in total. Observers identified 2,973 fishes, Reviewer A identified 2,993 and Reviewer B identified 2,976 (Table 12). Reviewer A had approximately 20 entries more than the observer and Reviewer B because Reviewer A recorded data for a group of sand dab flounder that were accidentally discarded en masse by the observer, while the observer and Reviewer B did not collect data for these.

Observer data included seven regulated species, three non-regulated species and one species group, Hake, not known, which encompassed *Urophycis*, *Merluccius* and *Physiculus* sp (including red, white and silver hake) (Table 12). Reviewer data included the same species and species groups as the observer, except Reviewer B data did not include four spot flounder. In addition, both reviewer data sets included three additional species groupings, which were flounder, not known for catch identified to the flounder level, groundfish, not known for catch identified no further than as a regulated groundfish species, and fish, not known for catch that could not be identified to any taxonomic level higher than a fish. These additional species groups accounted for 2.4% of Reviewer A and 16.7% of Reviewer B catch entries.

Table 12 Data entries from observer, Reviewer A and Reviewer B by species or species group (regulated species or groups that include only regulated species are marked with an asterisk)

Species/ Species Group	Observer Entries	Reviewer A Entries	Reviewer B Entries
Yellowtail flounder *	1264	1242	871
Sand dab flounder *	1161	1179	1150
Atlantic cod *	280	277	274
Winter flounder *	113	103	25
American plaice flounder *	95	61	100
Ocean pout *	34	34	35
Hake, silver	11	17	17
Hake, not known *	11	4	2
Witch flounder *	2	2	3
Four spot flounder	1	1	0
Monkfish	1	1	1
Flounder, not known	0	64	488
Fish, not known	0	7	9
Groundfish, not known *	0	1	1
Total	**2973**	**2993**	**2976**

The pairing of observer to each of the reviewer data sets resulted in a total 2,918 species matching pairs between observer and Reviewer A and 2,917 species matching pairs between observer and Reviewer B.

3.3.2. Comparisons

Individual Fish Identification by Observer and Reviewers

Catch pairs between observer and reviewer catch entries were compared for identification matches. Observer and Reviewer A comparisons for sand dab flounder and Atlantic cod had identical identification in 100% and 99% of observer entries respectively while yellowtail flounder and ocean pout were matched for 97% and winter flounder for 91% of observer entries. American plaice flounder was only matched for 63% of the observer entries (Table 13). Table 14 shows that 23 of the 29 non-matching American plaice flounder entries between observer and Reviewer A data were entered as flounder, not known, indicating that they could not be identified as a specific flounder species. Anecdotal information from the reviewer suggests that in many of these cases the reviewer was able to narrow identification down to American plaice flounder or yellowtail flounder but further identification was not possible because the mouth was not clearly visible, nor was the yellowtail flounder distinctive yellow colored ventral caudal peduncle area. A complete list of identification features used by reviewers, and the frequency with which they were used, are included as Appendix II.

Reviewer B comparisons to observer identification for Atlantic cod , sand dab flounder and ocean pout had a high proportion of matches with 99%, 98% and 94% observer entries matched respectively, which were within one and three percent points of the Reviewer A comparison results for these species. American plaice flounder comparisons for Reviewer B data also had similar results as Reviewer A with 66% of observer entries matched. Furthermore, Table 15 shows a similar distribution of un-matched American plaice observer entries with 18 out of 26 non-matched observer entries identified as flounder, not known by Reviewer B.

Reviewer B to observer identification comparisons had had very different results to Reviewer A to observer comparisons. While Reviewer A identification of yellowtail flounder and winter flounder matched 97% and 91% of observer entries for each species respectively, Reviewer B matched 66% of yellowtail flounder and 19% of winter flounder observer entries. Table 15 shows that 62 out of 88 (or 70%) of the non-matched winter flounder and 382 out of 423 (or 90%) of the non-matched yellowtail flounder were entered as flounder, not known.

Results from hake, not known, witch flounder, silver hake, and four spot flounder comparisons are inconclusive because these species had less than a dozen entries compared; however, the data suggested that reviewers had difficulties identifying hake, not known which is consistent with anecdotal information provided by reviewers that the specimens identified as fish, not known were likely to be very small hake, not known (~10cm) for which identifying characteristics were not discernible (Figure 11).

Table 13 Number of paired observer entries by species/species group with the corresponding number of reviewer identification matched entries (Percentages of paired observer entries matched by reviewer are shown for species/species groups with over 30 entries)

Species / Species Group	Reviewer A			Reviewer B		
	Paired Observer Entries	Reviewer ID Matches	%	Paired Observer Entries	Reviewer ID Matches	%
Yellowtail flounder	1253	1220	97%	1255	832	66%
Sand dab flounder	1150	1147	100%	1151	1124	98%
Atlantic cod	277	274	99%	275	272	99%
Winter flounder	110	100	91%	109	21	19%
American plaice flounder	78	49	63%	76	50	66%
Ocean pout	34	33	97%	34	32	94%
Hake, not known	10	4		11	2	
Witch flounder	2	2		2	2	
Silver Hake	3	2		3	3	
Four Spot flounder	1	1		1	0	
Total	**2918**	**2832**	**97%**	**2917**	**2338**	**80%**

Table 14 Catch identification matrix between observer and Reviewer A matched pairs

Observer \ Reviewer A	Atlantic cod	American plaice flounder	Four Spot flounder	Sand dab flounder	Winter flounder	Witch flounder	Yellowtail flounder	Hake, not known	Silver Hake	Ocean pout	Fish, not known	Flounder, not known	Groundfish, not known	Total
Atlantic cod	274						1			1			1	**277**
American plaice flounder		49					6					23		**78**
Four Spot flounder			1											**1**
Sand dab flounder				1147			2					1		**1150**
Winter flounder					100		3					7		**110**
Witch flounder						2								**2**
Yellowtail flounder		1		6			1220					26		**1253**
Hake, not known								4			6			**10**
Silver Hake									2		1			**3**
Ocean pout	1									33				**34**
Total	**275**	**50**	**1**	**1153**	**100**	**2**	**1232**	**4**	**2**	**34**	**7**	**57**	**1**	**2918**

Table 15 Catch identification matrix between observer and Reviewer B matched pairs

Observer \ Reviewer B	Atlantic cod	American plaice flounder	Sand dab flounder	Winter flounder	Witch flounder	Yellowtail flounder	Hake, not known	Silver Hake	Ocean pout	Fish, not known	Flounder, not known	Groundfish, not known	**Total**
Atlantic cod	272					1					1	1	**275**
American plaice flounder		50	1	1		6					18		**76**
Four Spot flounder					1								**1**
Sand dab flounder		1	1124			5					21		**1151**
Winter flounder		4	5	21		17					62		**109**
Witch flounder					2								**2**
Yellowtail flounder		32	7			832			2		382		**1255**
Hake, not known							2		1	8			**11**
Silver Hake								3					**3**
Ocean pout			2						32				**34**
Total	**272**	**87**	**1139**	**22**	**3**	**861**	**2**	**3**	**35**	**8**	**484**	**1**	**2917**

Species identification difficulties were not annotated into the data record by the reviewers in a standardized manner that would allow a quantitative analysis. However reviewer comments and post-review interviews revealed that the main factor that prevented discerning identifying characteristics in catch was the effect caused by the water pushing the fish down the discard chute. Sometimes the water flow would make the fish seem blurry or be strong enough to cause foam to form on the discard chute and cover parts of the fish. Flounders and smaller round fish were affected the most by this. For some flounders, species identification was not possible if the size and shape of the mouth was not visible to the reviewer. Reviewers also commented that increasing the resolution of the images would facilitate species identification.

Figure 11 Example images of fish where water flow obscured identifying characteristics (left) and where water flow did not have a detrimental effect on identification (right)

4. DISCUSSION

The results show that the methodology of using length-weight relationships to estimate the weight of regulated groundfish species using EM data deserves further investigation. Although the weights were statistically different, the differences were on average within 4% to 8% for flounders and within 8% to 12% Atlantic cod. Furthermore, there was evidence that using observer fish lengths and length-weight relationships to calculate weight is comparable to actual weights taken by observers, since there was no statistical difference between observer calculated weights and actual weights.

The preliminary estimates of the difference between observer weights and reviewer weights using length-weight relationships for flounders and Atlantic cod may be used to understand the potential impact of these differences in the context of estimating discard weights for a NE groundfish vessel. Throughout the experimental trips, Vessel A discarded approximately 49 pieces of Atlantic cod per trip which would represent a weight difference of -10 lbs to +15 lbs on average per trip (using Reviewer A and Reviewer B mean weight differences respectively). Assuming that the vessel does 80 trips in a fishing season, discarded weight from reviewer length measurement data would roughly represent a difference of -800 lbs to 1,200 lbs compared to observer actual weights. Vessel A discarded approximately 285 flounders per trip which would represent a weight difference of -6 lbs or -13lbs on average per trip and could translate to underestimating discarded catch by about 480lbs to 1,040lbs compared to observer actual weights over 80 trips in a fishing season.

It is worth noting that not all measurements were incorporated in the analysis. The estimated weight differences using length-weight relationships were based on instances where the reviewers were successful at obtaining a full length measurement of the fish. This would be an issue for calculating the total weight using reviewer lengths as the weight from these fish could be grossly underestimated. In an operational program these instances would have to be eliminated or their impact mitigated by, for example, applying sampling techniques using the complete length measurement data.

Instances when reviewers could not measure the full length of the fish were mostly due to partial images, which affected approximately 14% of the reviewer measurements. This issue could be minimized through a combination of changes to increase the success of having the entire length of the fish in the camera view. These changes could include: increasing the amount of frames per second recorded in the EM video, changing the catch handling process (such as slowing down the flow of the fish or holding the fish in camera view instead of sliding it) or modifying camera set-up to cover a larger area (in effect increasing the time the fish would be in camera view).

There were two main types of variation in the differences between reviewer and observer calculated weights using length-weight relationships: among reviewers and among fishing trips. Differences between reviewer estimates were particularly evident in Atlantic cod estimates, which were underestimated by Reviewer A and overestimated by

Reviewer B. It was not possible to identify the source of the difference although it was likely caused by differences in reviewer technique.

It is likely that differences between fishing trips were at least partially caused by inconsistency in the quality of the graduation marks, which varied over the course of the experiments. Furthermore, poor quality graduation marks could have been a factor affecting reviewer to observer comparisons.

Catch volume per species per haul was too low throughout the experiment to be able to apply a volumetric technique successfully. The mean weight differences using bushel baskets were relatively low (within two pounds for gadids and within six pounds for flounders) per basket (and hence per haul); however, the distribution of the differences had a wide spread and was not centered on the mean. For example, approximately one third of reviewer flounder weight estimates using bushel baskets overestimated the basket weight by 12 pounds. This was mostly due to over half of the basket fullness estimates being rounded up to ¼ full when they had four or less fish in them.

Using the small baskets reduced the weight difference between reviewer-estimated weight and actual weights (mean and spread) compared to the larger bushel baskets, even though with the smaller baskets there was greater disagreement estimating volume between ¼ and ½ fullness between observer and reviewers as well as amongst reviewers.

The exploratory experiment results for one of the reviewers established that it is possible for a reviewer to successfully identify discarded yellowtail flounder, Atlantic cod, sand dab flounder, winter flounder and ocean pout from EM data up to 91% to 100% of the time, based on the Reviewer A data set. However, there were differences between reviewers identification success. While Reviewer A was successful at identifying yellowtail flounder 97% and winter flounder 91% of the time, Reviewer B's success rate was 66% and 19% respectively. Furthermore, Reviewer A used the general species grouping flounder, not known for 2% of catch entries compared to 16% for Reviewer B; suggesting differences in success finding identifying characteristics on the video data. The differences in identification success rate between reviewers could be due to a combination of differences in experience identifying catch on video between reviewers and the characteristics selected by each reviewer. Fish identification through video often requires recognizing characteristics differently than an observer handling the fish would.

Consistent identification of American plaice flounder was difficult for both reviewers (63% and 66% matching identification for Reviewer A and Reviewer B respectively). A large mouth is one of the main identifying characteristics for distinguishing American plaice from yellowtail flounder. Reviewer A reported that the main issue preventing consistent identification of American plaice was that the video data did not clearly show the fish mouth due to foam in the water flow obscuring the fish mouth. This issue could be resolved by reducing the amount of water flowing when a discard chute is used, or modifying catch handling or equipment set-up to ensure that the fish mouth is visible in the video.

The data source pairing process was aimed at minimizing misalignment between the data sets. However, some comparison results may have been affected by errors innate in the comparison method itself rather than in measurement error or misidentification. In many cases, in particular for length data within a specific species or for species identification between similar species, these errors cannot be detected and hence quantified. In cases that include species identification matching pairs of species that are clearly different the alignment errors become more apparent. For example, the Atlantic cod to ocean pout comparisons between both reviewers and observer (likely recorded in different order) or hake, not known to ocean pout comparisons between Reviewer B and observer (likely the reviewer entering the wrong species name by mistake).

5. CONCLUSIONS

The objective of this work was to provide an initial assessment on whether these methodologies should be pursued further. The data collected during these exploratory experiments was sparse and the results presented in this report are preliminary. Additional data collection is needed in order to improve species identification and weight estimation, test the methodologies on a greater number of species and develop operational methods. Nevertheless, the preliminary results show that the use of length-weight relationships is a promising method for estimating discarded weight of regulated species and that the EM video can be used to consistently recognize indentifying characteristics on several species while others require more work. A volumetric methodology using bushel baskets is not appropriate for accurate weight estimation in applications where low volumes need to be estimated and more work is needed for evaluating whether weight estimates using smaller baskets may be adequate for estimating low catch volumes. Overall the results are positive, especially considering that there was limited opportunity to improve the methodologies since the experiments spanned only 14 trips, and we do not see any obvious obstacles to working on resolving the outstanding issues identified through further work.

Further work on evaluating these methodologies should involve two aspects: expanding the experiments and developing operational methods. Expanding the experiments is needed in order to compare observer and reviewer data for additional species and to improve the experimental design based on the results from the exploratory experiments. In particular, future work should target trips where hake species are expected given that this species group was highlighted as difficult to identify in the New England EM pilot 2010 report (Pria *et al.*, 2011) and the exploratory experiments included less than a dozen records.

Collecting the data required to ground truth these methodologies against observer data during fishing trips is difficult because it is dependent on the schedule and type of fishing the participating vessels are pursuing during experimental data collection. It may take a long time to collect the amount of data necessary for rigorous analysis. Alternatively, these methodologies may be tested in a laboratory environment with sample fish.

The second aspect that requires further work would be to use the lessons learned from the exploratory experiments in order to develop an operational methodology, which would have an on-board component (including equipment configuration as well as catch handling) and data analysis components. This aspect would need to take into account specific requirements for each gear type in the fishery (longline, gillnet and trawl).

An operational on-board methodology would continue to require a set-up where fish are presented to a close-up orthogonal camera view one-by-one for identification and measurement to allow the reviewer to measure and identify the discards. However, the specifics of the experimental observer on-board methodology would need to be adjusted so that captains and crew could carry it out within the operational reality of the vessel.

The reviewer methodology would also need to be adjusted towards maximizing operational efficiencies instead of the experiment's focus on collecting data in a way that allowed alignment to the observer data set on an individual fish or basket level. Other aspects of developing operational methods include training of reviewers and standardizing vessel set-ups to reduce variation in EM estimates.

These two aspects, the expansion of comparisons between observer and reviewer data and the development of operational methodologies, could not occur on the same vessel at the same time. The experiment expansion could take place initially or both aspects could occur in parallel, where some experimental data collection takes place strategically during the best data collection opportunities while other vessels take part in the operationalization of promising methodologies.

When weighing the need to expand the comparison between EM and reviewer data, a determination should be made in balancing rigorous scientific validation and operational realities. Although there was a statistically significant difference between reviewer and observer calculated weights, it is important to assess whether, in the event that this methodology was used in an operational program, this difference would constitute an acceptable risk or whether it needs to be reduced and, if so, to what level. Additionally, when considering the risk associated with using these methodologies for providing weight estimates by species from EM data based on comparisons to at-sea observer data, it is important to frame the issue in the context that there would be measurement and data collection errors intrinsic in any data collection method, including EM and human observer data.

6. ACKNOWLEDGEMENTS

These exploratory experiments were funded by NMFS as part of the New England EM pilot project. We are grateful for the cooperation, support and patience of Captain Frank Mirarchi and Captain Gary Libby and their respective crew, who participated in the experiments. We thank Amy Van Atten and Nichole Rossi for their advice and assistance during the experiments. We are also appreciative of the dedication of Kelly Neville and Glenn Chamberlain from the FSB who observed the experimental trips and reviewed EM data, as well as provided insight throughout the planning and execution of the experiments. Mark Hager from East West Technical Services provided field support as needed throughout the experiments. Several Archipelago staff participated in this project including: Kim Astle, Steven Roias, and Kathryn Archibald.

REFERENCES

Archipelago Marine Research Ltd. 2012. New England Electronic Monitoring Project In-season Data Summary January 1, 2012 to March 31, 2012. Unpublished report prepared for the Northeast Fisheries Science Center Fisheries Sampling Branch by Archipelago Marine Research Ltd, Victoria, BC Canada. 26 p.

Chase, P.D. and Galbraith J.K. 2004. Observer Training: An Overview of Field Identification Techniques - Gulf of Maine to Cape Hatteras, NC. NOAA Northeast Fisheries Science Center, Ecosystems Surveys Branch. 68p.

Crowley, P. H. 1992. Resampling Methods for Computation-Intensive Data Analysis in Ecology and Evolution. Annual Review of Ecology and Systematics. Vol. 23. pp. 405-447

Douglas, J., Robins, C. R. and Peterson, R. T. and Ray, C. 1999. A field guide to Atlantic Coast Fishes of North America. New York, NY: Houghton Mifflin Harcourt Publishing.

Froese, R. and Pauly D. Editors. 2012.FishBase. World Wide Web electronic publication. www.fishbase.org.

Gilbert, C. R. and Williams, J. D. 1993. National Audubon Society Field guide to Fishes. North America. New York, NY: Alfred A. Knopf.

Pria, M. J., Bryan, J. and McElderry, H. 2011.New England Electronic Monitoring Project 2010 Annual Report. Unpublished report prepared for the Fisheries Sampling Branch of the Northeast Science Centre by Archipelago Marine Research Ltd., Victoria BC Canada. 69 p.

Wigley, S. E., McBride, H. M. and McHugh, N. J. 2003. Length-Weight Relationships for 74 Fish Species Collected during NEFSC Research Vessel Bottom Trawl Surveys, 1992-99. NOAA Technical Memorandum NMFS-NE-171. 26 p.

APPENDIX I – STATISTICAL TESTS LENGTH EXPERIMENT RESULTS

Reviewer A and Observer Estimated Weight Differences-Linear Regression Model

```
lm(formula = Weight Difference ~ Species + tripID)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.57599 -0.03146  0.00272  0.04091  1.36163 

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)    
(Intercept)                   0.018231   0.026422   0.690 0.490344    
Species.Atlantic cod         -0.124483   0.027344  -4.552 5.91e-06 ***
Species.Atlantic halibut     -0.022509   0.123345  -0.182 0.855233    
Species.haddock              -0.052480   0.088962  -0.590 0.555370    
Species.ocean pout           -0.084951   0.089643  -0.948 0.343515    
Species.sand dab flounder     0.006761   0.024529   0.276 0.782872    
Species.winter flounder       0.019720   0.033242   0.593 0.553165    
Species.witch flounder       -0.028773   0.173028  -0.166 0.867958    
Species.yellowtail flounder -0.012136   0.024387  -0.498 0.618844    
tripID.311820.04             -0.020215   0.021377  -0.946 0.344533    
tripID.311821.01              0.014264   0.019463   0.733 0.463775    
tripID.311822.03             -0.074205   0.020465  -3.626 0.000301 ***
tripID.311823.02             -0.073174   0.020429  -3.582 0.000356 ***
tripID.311827.01             -0.084007   0.028228  -2.976 0.002986 ** 
tripID.311828.01             -0.170029   0.024515  -6.936 6.96e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1713 on 1076 degrees of freedom
Multiple R-squared: 0.2352,     Adjusted R-squared: 0.2253 
F-statistic: 23.64 on 14 and 1076 DF,  p-value: < 2.2e-16
```

Reviewer B and Observer Estimated Weight Differences- Linear Regression Model

```
lm(formula = Weight Difference ~ Species + tripID)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.22049 -0.03001  0.00663  0.03485  0.51322 

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)    
(Intercept)                  -0.076962   0.014792  -5.203 2.38e-07 ***
Species.Atlantic cod          0.283879   0.018819  15.085  < 2e-16 ***
Species.Atlantic halibut     -0.026634   0.096931  -0.275   0.7835    
Species.haddock               0.020637   0.096873   0.213   0.8313    
Species.ocean pout            0.031057   0.051367   0.605   0.5456    
Species.sand dab flounder     0.011860   0.013599   0.872   0.3833    
Species.winter flounder       0.010175   0.018638   0.546   0.5852    
Species.witch flounder        0.013792   0.096931   0.142   0.8869    
Species.yellowtail flounder   0.011347   0.013606   0.834   0.4045    
tripID.311820.04              0.073080   0.012404   5.891 5.24e-09 ***
tripID.311821.01             -0.046200   0.010561  -4.375 1.34e-05 ***
tripID.311822.03              0.129755   0.011729  11.063  < 2e-16 ***
tripID.311823.02              0.024867   0.011517   2.159   0.0311 *  
tripID.311827.01              0.002937   0.015215   0.193   0.8470    
tripID.311828.01              0.025216   0.019597   1.287   0.1985    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1 

Residual standard error: 0.09595 on 990 degrees of freedom
Multiple R-squared: 0.4972,     Adjusted R-squared: 0.4901 
F-statistic: 69.94 on 14 and 990 DF,  p-value: < 2.2e-16
```

APPENDIX II –SPECIES IDENTIFICATION FEATURES USED BY REVIEWERS

Please note that the feature descriptions were not standardized across reviewers. As a result, descriptions between reviewers may overlap in situations when one reviewer described a feature in slightly more or less detail than the other reviewer or used a combination of features as one.

Feature	Reviewer A		Reviewer B	
	Times Used	% Used	Times Used	% Used
American plaice				
Large Mouth	61	100%	75	75%
Right Eyed	61	100%	22	22%
Convex Tail	15	25%	68	68%
Narrow Caudal Peduncle	43	70%		
Slender Body Profile with Round/Spade Shaped Tail	1	2%		
Large Mouth			1	1%
Thick Body			1	1%
Lack of Other Flounder Characteristics			16	16%
Total entries for American plaice	**61**		**100**	
Atlantic cod				
White lateral line	271	97%	268	98%
Three dorsal fins	236	85%	212	78%
Coloration	209	75%	67	25%
Sub-terminal mouth	90	32%	2	1%
Chin barbel	12	4%		
Slightly forked/squared tail	4	1%		
Two anal fins	1	0%		
Large eyes	1	0%		
Total entries for Atlantic cod	**278**		**273**	
Fish, not known				
Slender body	5	71%		
Long and Silver	1	14%		
No identifying characteristic recorded			9	100%
Total entries for fish, not known	**7**		**9**	
Flounder, not known				
Right eyed	62	97%		
Narrow caudal peduncle	33	52%		
Small mouth	15	23%		
Slender body profile and round tail	6	9%		
Slender body profile	4	6%		
Dark dorsal surface	3	5%		
Thick caudal peduncle	2	3%		
Left eyed	1	2%		
Large mouth	1	2%		
Flat body shape			474	97%
Round body shape			1	0%
No identifying characteristic recorded			15	3%
Total entries for flounder, not known	**64**		**490**	

Feature	Reviewer A		Reviewer B	
	Times Used	% Used	Times Used	% Used
Fourspot flounder				
Left eyed	1	100%		
Large Mouth on Slender Body	1	100%		
Spade-shaped caudal fin	1	100%		
Total entries for fourspot flounder	**1**			
Groundfish, not known				
Three dorsal fins	1	100%		
Sub-terminal Mouth	1	100%		
Mottled brown body color	1	100%		
Round body shape			1	100%
Total entries for Groundfish, not known	**1**		**1**	
Hake, not known				
Long second dorsal fin	4	100%	2	100%
Long anal fin	2	50%		
Round caudal fin	2	50%		
Long pelvic fin			2	100%
Total entries for hake, not known	**4**		**2**	
Monkfish				
Large head with huge mouth followed by short tapering body	1	100%	1	100%
Small fleshy pelvic fins posterior to pectoral fins	1	100%	1	100%
Large body to tail ratio			1	100%
Distinctive fins			1	100%
Total entries for monkfish	**1**		**1**	
Ocean pout				
Very elongate body with reduced caudal fin	34	100%	15	43%
Large orange/yellow pectoral fins	32	94%		
Large fleshy mouth	18	53%		
Dorsal fin ends well before tail	6	18%		
Reduced tail	2	6%		
Large pectoral fins	1	3%		
Orange/brown body color	1	3%		
Rounded pectoral fin			19	54%
Continuous anal/caudal fin			28	80%
Continuous dorsal fin			7	20%
Total entries for ocean pout	**34**		**35**	
Sand dab flounder				
Very round body profile	1182	100%	1151	100%
Left eyed	1169	99%	8	1%
Heavy Spotting on Fins	1143	97%	1138	99%
Visible gut cavity	4	0%		
Large mouth			12	1%
Convex tail			2	0%
Total entries for sand dab flounder	**1182**		**1152**	

Feature	Reviewer A		Reviewer B	
	Times Used	% Used	Times Used	% Used
Silver Hake				
Long second dorsal fin	17	100%		
Long anal fin	9	53%		
Round caudal fin	1	6%		
Coloration			17	100%
Large mouth			7	41%
No barbel			1	6%
Total entries for silver hake	**17**		**17**	
Winter flounder				
Right eyed	102	99%	2	8%
Small mouth	70	68%	20	80%
Thick caudal peduncle	103	100%		
White ventral surface (opaque)	32	31%		
Thick body			25	100%
Flat lateral line			1	4%
Convex tail			3	12%
No upturned mouth			1	4%
Total entries for winter flounder	**103**		**25**	
Witch flounder				
Right eyed	2	100%		
Small mouth	2	100%		
Concave pelvic region	1	50%		
Narrow caudal peduncle with round tail	1	50%		
Dark around anal fins			2	67%
Dark spot on pectoral fin			2	67%
Thin body			2	67%
Total entries for witch flounder	**2**		**3**	
Yellowtail flounder				
Right eyed	1242	100%	163	19%
Upturned mouth/snout	1173	94%	866	100%
Dirty yellow ventral surface of caudal peduncle	338	27%	45	5%
Small mouth	1175	94%		
Narrow caudal peduncle	944	76%		
Slender body with round tail	3	0%		
Convex tail			807	93%
Large fleshy lip			18	2%
Total entries for yellowtail flounder	**1242**		**869**	