Final Report

CONTINUING THE TIME SERIES: CALIBRATING THE NMFS SEA SCALLOP SURVEY TO THE R/V *HUGH R. SHARP*

Award Number: 09-SCA-07

Submitted to:

National Marine Fisheries Service
Northeast Fisheries Science Center
Cooperative Research Program
166 Water Street
Woods Hole, Massachusetts 02543-1026

Submitted by:

David B. Rudders
William D. DuPaul

Virginia Institute of Marine Science
College of William and Mary
Gloucester Point, Virginia 23062

VIMS Marine Resource Report No. 2010-4

May 28, 2010

Project Summary

The annual synoptic survey of the United States sea scallop resource by the National Marine Fisheries Service (NMFS) represents a vital component of the information used to manage the fishery. Sea scallop abundance indices obtained from this survey have been generated from research cruises aboard the R/V *Albatross IV* since the 1970s. In addition to the continuity of vessel platform, the survey dredge had also been consistent throughout the time series. Research vessels have a finite life span and improvements to sampling gear are sometimes required. Care, however, must be taken to account for any changes in catchability that might occur due to altering a vessel or sampling gear. Systematic error may be introduced into the time series if the indices are not adjusted to account for these changes the sampling protocol.

The summer of 2007 represented the final year of operations for the R/V *Albatross IV.* In anticipation of the retirement of this vessel, the Virginia Institute of Marine Science (VIMS) in conjunction with the Northeast Fisheries Science Center (NEFSC) and the sea scallop industry conducted vessel calibration experiments during the 2007 NMFS sea scallop survey. These experiments, conducted aboard two commercial sea scallop vessels, were intended to preserve the continuity of the time series by providing fishing power correction (FPC) factors relative to the R/V *Albatross IV*. This information would facilitate the use of the calibrated commercial vessels to conduct the survey, or at least form a link from the R/V *Albatross IV* to any future survey platform. In addition to calibrating two potential vessel platforms, an updated dredge design (developed by the Sea Scallop Survey Advisory Panel (SSSAP)) was also used in the experiment. The new dredge design, towed simultaneously with the standard dredge was used to anticipate and account for a potential change in survey gear. In total FPC factors were estimated for four different vessel-gear configurations with respect to the catch per unit effort (CPUE) of sea scallops. Correction factors were generally small, indicating that the systematic bias associated with different vessel-gear configurations was not large, although significant scallop catch at length differences were observed for some vessel/gear combinations. In particular, one vessel used in the study (F/V *Nordic Pride* towing the standard dredge design) performed virtually equivalently to the R/V *Albatross IV* (1.05 statistically non-significant) with respect to total catch of sea scallops. This vessel/gear combination was considered to represent the NMFS sea scallop time series.

Subsequent to the completion of the 2007 calibration experiments, the decision was made to utilize the R/V *Hugh Sharp* as the survey platform for the NMFS sea scallop survey. As a result of the R/V *Albatross IV* being unavailable for direct calibration studies, the 2007 VIMS experiments represents a major link to the historic time series. In 2009, we conducted an additional calibration experiment utilizing the F/V *Nordic Pride* in an effort to calibrate the R/V *Hugh Sharp* to the existing time series. Results indicate that the R/V *Hugh Sharp* was slightly more efficient (FPC ~1.10 statistically non-significant) relative to the F/V Nordic Pride with evidence to support length based differences in catchability. Based on the results from the pooled over length catches, there is no evidence to support significant differences in catchability between the vessels (R/V *Albatross IV*, F/V *Nordic Pride and R/V Hugh Sharp*) and no strong evidence to suggest the imposition of an FPC is warranted.

Project Background

Fishery surveys provide information that is vital for the assessment of aquatic resources. Information supplied by annual synoptic surveys of fish and shellfish stocks serves a variety of important roles. Indices of abundance generated by surveys, track relative changes in population abundance over time, and depending on the configuration of the gear used, the presence and relative magnitude of recruitment events. Surveys can provide information to detect changes in species assemblage over time, as well as providing samples to assess changes on an organismal level (Hilborn and Walters 1992; Gunderson 1993). Perhaps most important, the information gathered by annual fishery surveys populate stock assessment models. These models, in turn, estimate critical components of the assessed stock such as estimates of present and future abundance, as well as mortality rates. With these estimates, guidance to managers relating to responsible levels of harvest can be supplied in order to achieve management goals. Given the importance of the time series to both stock assessments and ultimately the responsible and effective management of marine resources, the onus lies on maintaining a high level of long-term data quality. It is essential to preserve the continuity of the time series and is vital to insure its utility as a source of information in both retrospective as well a forward projecting modeling efforts.

When monitoring relative changes in abundance over time through annual fishery surveys, the implicit assumption in comparing the results between years is that the measured index of abundance is proportional to the actual abundance. The proportionality constant known as the catchability coefficient (q) is assumed in the strictest sense to be constant, or at least stationary (varying without trend) (Kimura and Somerton 2006). To satisfy the assumption of stationarity of q, researchers must standardize all components of the survey methodology. Should changes to the methodology occur, it is vital to calibrate the new methodology to the old to ensure comparability to existing time series. In the event of changes to survey protocols,

calibration experiments allow for the utilization of the entire time series to seamlessly be included in stock assessment models.

Many components of fishery surveys can be standardized through time to satisfy the assumption of stationarity of the catchability coefficient. Maintaining a standard survey design, fishing gear and sampling methodology are excellent practices, however a major impetus necessitating calibration studies is either the replacement of a dedicated survey vessel, the utilization of multiple vessels to complete a given survey or changes to the survey gear (Tyson *et. al.,* 2006). Differences in survey vessels can have a profound effect on the magnitude of the CPUE observed during in a given survey. This vessel effect has the potential to introduce bias into the time series if left unaccounted for (Pelletier, 1998). Calibration experiments designed to quantify the relative differences in fishing power can account for any changes to the survey methodology (vessel, gear, design, etc.) and are used to adjust the time series moving forward (von Szalay and Brown, 2001).

The methodology for conducting fishing vessel inter-calibration experiments was reviewed by Pelletier (1998). Pelletier observed that these experiments generally fall into two experimental design categories. The first design was an independent haul approach which sampled in a confined area with the assumption of uniform fish abundance and environmental conditions throughout the area. Experiments utilizing this approach generally estimated the FPC factors within a randomized block analysis of variance (ANOVA) framework with each tow representing a block. In general, this design introduces considerable spatial and temporal variability. The result of this variability is a requirement of a large effective sample size to detect differences in the block-treatment effect. Additionally, these additional sources of unaccounted variability have the potential to affect the precision of estimated FPC factors (Pelletier, 1998). The second and much more common experimental approach was the paired design, where two vessels occupied tows either simultaneously separated by a safe, but small distance, or

reoccupied the same tow path in close succession. This paired design has the advantage of reducing the spatial and temporal variability relative to the independent haul method.

Regardless of the survey design used, FPC factors lend themselves to certain classes of analytical approaches. First used by Robson (1966) variations of log–transformed multiplicative models have been a common analytical approach (Sissenwine and Bowman 1978; Wilderbuer *et al.* 1998). Another approach involves a ratio estimator of the mean CPUE of the two gears or vessel/gear combination (Wilderbuer *et al.* 1998; Tyson *et al.* 2006). These two analytical approaches are sensitive to implicit assumptions relating to the availability of fish in the tow path. Violation of these assumptions is possible due to the nature of some habitats sampled as well as the contagious distribution of fish (Lewy *et al.* 2004).

Kappenman (1992) developed an approach to estimate relative fishing power based upon a ratio of scale parameters for two positive random variables (CPUE). The underlying assumption of this method is that the two CPUE distributions for a given species have the same underlying shape, but different scales. With this technique, a FPC factor is estimated from the ratio of the two scale parameters. This approach is attractive relative to more traditional analytical procedures (randomized block ANOVA, ratio of mean CPUE and least squares regression) due to the lack of assumptions required. The Kappenman technique does not require a strict pairing of tows and there is no assumption of equal fish density available for each tow. Utilizing the same data set, Wilderbuer *et al.* (1998) compared four approaches (randomized block ANOVA, ratio of mean CPUE and least squares regression, Kappenman) and found similar and superior performance for the randomized block ANOVA and Kappenman. While procedures for calculating 95% confidence intervals exist for randomized block ANOVA, ratio of mean CPUE and least squares regression, one does not exist for the Kappenman estimator. von Szalay and Brown (2001) used a bootstrapping approach to resample the CPUE data from the two vessels and estimate the variance of the Kappenman estimator.

More recently, FPCs have been estimated with analytical approaches utilizing generalized linear models (GLM) and generalized linear mixed models (GLMM) (Helser *et al.* 2004; Lewy *et al.* 2004, Cadigan *et. al.*, 2006). In addition to estimating FPC, Lewy *et al.* (2004) was able to estimate the disturbance effect that occurs when two vessels consecutively tow along similar tow paths causing a change in the availability of fish. Both the Kappenman method (1992) as well as the GLMM approach by Helser *et al.* (2004) have been used to examine FPCs in surveys where multiple vessels have been used in a given survey. These studies are interesting in the fact that explicit calibration experiments were not performed, yet survey results were analyzed *a posteriori* and allowed the consolidation of multiple data sets data into calibrated indices of abundance (von Szalay and Brown, 2001; Helser *et al.* 2004). This approach can have benefits in reducing the inherent spatial and temporal variability when large geographic areas are surveyed especially with highly mobile or migratory species.

Regardless of the technique used to estimate a FPC, the critical decision is whether to apply the correction to the existing time series. Traditionally, 95% confidence intervals were use to decide whether to apply the factor. If the interval spanned one (implying there was no difference in vessel/gear variant for a given species) a correction was not applied. Conversely, if the interval did not include one then the correction was applied. This thinking can be problematic in the sense that FPCs are notoriously imprecise (i.e. wide confidence intervals that include unity) and true differences in relative fishing power may be incorrectly rejected. Munro (1998) developed an objective decision rule for the application of the correction factor based on the conjecture that the application of a FPC was only beneficial if it reduces the error in the estimate of the mean CPUE.

With the replacement of the R/V *Albatross IV* by the R/V *Hugh Sharp* as the vessel platform for the sea scallop survey, VIMS in conjunction with the NEFSC and the sea scallop industry conducted a vessel calibration experiment during the 2009 NMFS sea scallop survey. The rationale for this experiment was based on the results from our 2007 experiments and

sought to link the existing time series by determining FPC factors of a commercial vessel (F/V *Nordic Pride*) that was demonstrated to perform equivalently to the R/V *Albatross IV* to the new vessel platform. We conducted paired tows aboard the same vessel and with the same gear as in a portion of our 2007 study. The objective of this experiment was to preserve the continuity of the time series by providing FPC factors of the R/V *Hugh Sharp* relative to the F/V *Nordic Pride* and indirectly calibrate the new vessel platform to the R/V *Albatross IV*.

Methods

The main objective of this study was to calibrate the new vessel platform used in the NMFS sea scallop survey. The sea scallop time series was represented by the F/V *Nordic Pride*, a vessel that was calibrated to the R/V *Albatross IV* in 2007. Implicit to this objective is the assessment of potential differences in relative efficiency between the vessel that represents the time series and replacement. In addition to differences in relative efficiency that result from a change in vessel platform, a re-designed scallop survey dredge was also assessed. Data from paired tows were used to quantify potential differences in vessel/gear combinations. These paired tows compared the CPUE of sea scallops between a commercial vessel (F/V *Nordic Pride*) and the R/V *Hugh Sharp* on the mid-Atlantic leg of the 2009 NMFS sea scallop survey. The paired tows for the comparison were a subset of the 2009 survey effort, but were selected to representative of the domain sampled by NMFS. In addition to the geography sampled, the comparative tows were also representative of the range of biotic factors (scallop size, density and co-occurring species) as well as abiotic factors (depths, currents, and substrate) encountered.

Experimental Design

For this experiment, the commercial vessel was selected based on vessel characteristics. To be a candidate vessel to conduct offshore survey work, this vessel needed

to be able to sample in all portions of the sea scallops range. In the fishery, there is a wide range of vessels and not all vessels can effectively operate in all areas due to different prevailing weather and oceanic characteristics. The vessels also needed to be large enough to accommodate the scientific party as well as the vessel crew with ample space for the completion of sampling. Characteristics of vessels used in the study are shown in Table 1.

The calibration experiment was conducted within the context of the mid-Atlantic portion of the 2009 NMFS annual sea scallop survey (Figure 1). This survey utilizes a stratified random design to sample throughout the entire U.S. range of the sea scallop. (Serchuk and Wigley 1986). Due to regional differences in the composition of the substrate as well as hydrographic conditions, our goal was to sample throughout the geographic region sampled during this portion of the survey. Sampling cruises occurred during the first leg of the NMFS survey. This cruise occupied stations in the mid-Atlantic region, specifically the DelMarVa, Elephant Trunk and Hudson Canyon closed areas (Figure 2).

Sampling gear consisted of two sea scallop survey dredges towed simultaneously. The first dredge was the standard NMFS sea scallop survey dredge that has been in service, virtually unmodified since the 1970s. This dredge is 8' in width, with a dredge bag consisting of 2" rings. The twine top is comprised of 3.5" diamond mesh and there is a 1.5" liner throughout the dredge bag. There were no turtle excluder chains on this dredge. The second dredge used in this study was a modified version of the standard dredge developed by the SSSAP. In this document, this dredge will be referred to as the "prototype" dredge. The components of the prototype dredge are almost identical to the standard dredge (i.e. ring size, liner mesh size and twine top mesh size). Differences exist in relation to a slightly modified dredge frame, modifications to the ring bag and slight modifications to the mesh counts of the liner and twine top. A major difference between the standard and prototype dredge configurations is the addition of a wheel on the frame of the dredge as well as turtle/rock chains. The rationale behind the inclusion of chains for this dredge was to construct a dredge that was functional in all

areas sampled as well as being proactive in taking measures relating to the exclusion of sea turtles from sea scallop dredges.

While at sea, the sampling protocol included the re-occupation of sampling stations occupied by the R/V *Hugh Sharp.* Start/stop locations for each tow completed by the R/V *Hugh Sharp* were relayed to the commercial vessel via VHF radio. With the goal of re-occupying the stations as quickly as possible, a subset of stations was selected for re-sampling (the R/V *Hugh Sharp* conducts 24 hour operations, while the F/V *Nordic Pride* in this study sampled for roughly 16-18 hours/day). During the execution of the tow, the captain of the F/V *Nordic Pride* attempted to mirror the start/stop locations as close as possible. While it is safe to assume that there was some crossing of tow paths, it is unlikely that the tow path was duplicated precisely. For each comparative tow, the dredges were fished for 15 minutes with a towing speed of approximately 3.8-4.0 kts. High-resolution navigational logging equipment was used to accurately determine vessel position and speed over ground. Time stamps from the navigational log in conjunction with the tow level information recorded on the bridge were used to determine the location, duration and area fished by the dredges.

Sampling of the catch was conducted in the same manner as established by DuPaul *et. al,.*1989. For each paired tow, the entire scallop catch was placed in baskets. A fraction of these baskets were measured and subsequently expanded to estimate length frequency for the entire catch. The shell height of each scallop in the sampled fraction was measured in 5 mm intervals. This protocol allowed for the determination of the size frequency of the entire catch by expanding the catch at each shell height by the fraction of total number of baskets sampled. Finfish and invertebrate bycatch were quantified, with finfish being sorted by species and measured to the nearest 1 cm. Sampling protocol was similar on the R/V *Hugh Sharp*.

The standard data sheets, used since the 1998 Georges Bank industry-based survey, were used. The bridge log maintained by the captain/mate recorded location, time, tow-time

(break-set/haul-back), tow speed, water depth, catch, bearing, weather and comments relative to the quality of the tow. The deck log maintained by the scientific personnel recorded detailed catch information on scallops, finfish, invertebrates and trash.

Statistical Models

Scallop catch data from the paired tows provided the information to estimate differences in the fishing power of each vessel/gear combination tested. Assume that each vessel/gear combination tested in this experiment has a unique catchability. Let q_r equal the catchability of the R/V *Albatross IV* and q_f equal the catchability of the commercial vessel (F/V *Nordic Pride*) used in the study. The efficiency of the research vessel relative to the commercial vessel will be equivalent to the ratio of the two catchabilities.

$$\rho_l = \frac{q_r}{q_f} \qquad (1)$$

The catchabilities of each the vessel/gear combination are not measured directly. However, within the context of the paired design, assuming that spatial heterogeneity in scallop density is minimized, observed differences in scallop catch for each vessel will reflect differences in the catchabilities of the vessel/gear combinations tested. Our analysis of the efficiency of the research vessel relative to the commercial vessels consisted of two levels of examination. The first analysis consisted of an examination of potential differences in the total scallop catch per tow. Subsequent analyses investigate whether scallop size was a significant factor affecting relative efficiency. Each analysis incorporates an approach to account for within-tow variation in the spatial heterogeneity of scallop density.

Let C_{iv} represent the scallop catch at station *i* by vessel *v*, where v=r denotes the research vessel (R/V *Hugh Sharp*) and v=f denotes the commercial vessel (F/V *Nordic Pride*).

Let λ_{ir} represent the standardized scallop density for the ith station by the R/V *Hugh Sharp* and λ_{if} the standardized scallop density encountered by the F/V *Nordic Pride* commercial vessel. We assume that due to the tow paths taken by the respective vessels at tow *i*, the densities encountered by the two vessels may vary as a result of small-scale spatial heterogeneity as reflected by the relationship between scallop patch size and coverage by a standardized tow. The standardized unit of effort is a survey tow of 15 minutes at 3.8 kts. which covers a linear distance of approximately .95 nautical miles. The probability that a scallop is captured during a standardized tow is given as q_r and q_f. These probabilities can be different for each vessel, but are expected to be constant across stations. Assuming that catch by the F/V *Nordic Pride* is a Poisson process with mean equal to variance, the expected catch is given by:

$$E(C_{if}) = q_f \lambda_{if} = \mu_i \qquad (2)$$

Assuming the catch by the R/V *Hugh Sharp* is also a Poisson random variable the expected catch for this vessel is:

$$E(C_{ir}) = q_r \lambda_{ir} = \rho \mu_i \exp(\delta_i) \qquad (3)$$

Where $\delta_i = \log(\lambda_{ir} / \lambda_{if})$. For each station, if the standardized density of scallops encountered by both vessels is the same, then $\delta_i = 0$.

If the vessels encounter the same scallop density for a given tow, (i.e. $\lambda_{ir} = \lambda_{if}$), then ρ can be estimated via a Poisson GLM. This approach, however, can be complicated especially if there are large numbers of stations and scallop lengths (Cadigan *et. al*., 2006). The preferred approach is to use the conditional distribution of the catch by the research vessel at station i, given the total non-zero catch of both vessels at that station. Let c_i represent the observed value of the total catch. The conditional distribution of C_{ir} given $C_i = c_i$ is binomial with:

$$\Pr\left(C_{ic} = x \middle| C_i = c_i\right) = \binom{c_i}{x} p^x (1-p)^{r_i - x} \qquad (4)$$

Where p=ρ/(1+ρ) is the probability a scallop is captured by the research vessel. In this approach, the only unknown parameters is ρ and the requirement to estimate μ for each station is eliminated as would be required in the direct GLM approach (equations 2 & 3). For the Binomial distribution $E(C_{ir})=c_i p$ and $Var(C_i r)=c_i p/(1-p)$. Therefore:

$$\log\left(\frac{p}{1-p}\right) = \log(\rho) = \beta \qquad (5)$$

The model in equation 5, however does not account for spatial heterogeneity in the densities encountered by the two vessels for a given tow. If such heterogeneity does exist then the model becomes:

$$\log\left(\frac{p}{1-p}\right) = \beta + \delta_i \qquad (6)$$

Where δ_i is assumed to be normally distributed with a mean=0 and variance=σ^2. This model represent the formulation to estimate the vessel effect ($\exp(\beta_0)$) when scallop catch per tow is pooled over length.

Often, the replacement of a survey vessel presents an opportunity to make changes to the survey fishing gear. In those instances, the potential exists for the catchability of scallops at length, *l* to vary. Even in cases where the survey fishing gear remains the same, length effects are possible. Models to describe length effects are extensions of the models in the previous

section to describe the total scallop catch per tow. Again, assuming that between-pair differences in standardized scallop density exist, a binomial logistic regression GLMM model to reflect the situation where one vessel encounters more scallops, but they are of the same length distribution would be:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \delta_i + \beta_1 l, \delta_i \sim N(0,\sigma^2), i = 1,\ldots,n. \qquad (7)$$

In this model, the intercept (β_0) is allowed to vary randomly with respect to station.

The potential exists, however, that there will be variability in both the number as well as the length distributions of scallops encountered within a tow pair. In this situation, a random effects model that allows both the intercepts (δ_0) and slopes (δ_1) to vary randomly between tows is appropriate (Cadigan and Dowden, 2009). This model is given below:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \delta_{i0} + (\beta_1 + \delta_{i1})l, \delta_{ij} \sim N(0,\sigma_j^2), i = 1,\ldots,n,\ j = 0,1. \qquad (8)$$

Adjustments for sub-sampling of the catch and differences in area swept

Additional adjustments to the models were required to account for sub-sampling of the catch as well as differences in the observed area swept by the two gears. In some instances, due to high volume, catches for particular tows were sub-sampled. Often this is accomplished by randomly selecting a subset of the total catch (in baskets) for length frequency analysis. One approach to accounting for this practice is to use the expanded catches. For example, if half of the total catch was measured for length frequency, multiplying the observed catch by two would result in an estimate of the total catch at length for the tow. This approach would artificially

overinflate the sample size resulting in an underestimate of the variance, increasing the chances of spurious statistical inference (Millar *et. al.*, 2004; Holst and Revill, 2009). In our experiment, the proportion sub-sampled was consistent throughout each tow and did not vary with respect to scallop length. While experimental protocol dictates a standardized tow of roughly .95 nautical miles (3.8 kts. For 15 minutes), in practice variability exists in the actual tow distances covered by each vessel. These differences must be accounted for in the analysis to ensure that common units of effort are compared.

Let q_{ir} equal the sub-sampling fraction at station *i* for the vessel *r* and let d_{ir} be the areal coverage at station *i*, for vessel *r.* This adjustment results in a modification to the logistic regression model:

$$\log\left(\frac{p_i}{1+p_i}\right) = \beta_0 + \delta_{i0} + (\beta_1 + \delta_{i1})l_i + \log\left(\frac{q_{ir}d_{ir}}{q_{if}d_{if}}\right), \delta_{ij} \sim N(0, \sigma_j^2), i = 1,\ldots,n,\ j = 0,1. \quad (9)$$

The last term in the model represents an offset in the logistic regression (Littell, *et. al., 2006*).

We used SAS/STAT® PROC NLMIXED to fit the generalized linear mixed effects models. This procedure fits the specified nonlinear model by maximizing a approximation to the likelihood integrated over the random effects (Wolfinger, 1999). While PROC NLMIXED is limited in the number and complexity of the random effects that can be effectively modeled, Cadigan and Dowden (2009) reported that simulation results using this software were more reliable relative to procedures that relied on different approximation techniques.

Results

The comparative fishing experiment was conducted during the mid-Atlantic leg of the annual NMFS sea scallop dredge survey. The two vessels utilized during this study were the F/V *Nordic Pride* hailing from New Bedford, MA and the R/V *Hugh Sharp* from Lewes, DE (see Table 1 for vessel characteristics and Figure 2 for a map of the occupied stations). This cruise

occurred from May 9, 2009 to May 17, 2009. During that time 106 comparative tows were completed, with the eventual number of valid hauls being slightly lower due to a couple of fouled hauls. In general, the comparative stations were completed in the DelMarVa region of the mid-Atlantic Bight in both open access as well as areas currently managed under a rotational area management strategy. Included within the broad geographic range encompassed by the sampling, were the high density areas of the Elephant Trunk, DelMarVa and Hudson Canyon closed areas as well as open areas adjacent to the aforementioned rotational closed areas. The areas sampled resulted in comparative tows that included a wide range of scallop densities as well as the spectrum of depths and other biotic and abiotic factors generally encountered throughout this survey.

For this cruise, trip level information is shown in Table 2. Each tow pair was individually examined to determine whether it was appropriate to include in the final analysis. Gear related problems including hangs, flips and gear damage justified exclusion. In addition, a tow track based on the navigational information logged during the deployment was plotted and examined for spatial congruence. The relative tow locations of all tow pairs were consistent with the experimental protocol and no tow pairs were excluded based on this criterion. While the proximity of each tow within a given pair was acceptable, differences existed in the relative tow distances covered by the two vessels. In general, the R/V *Hugh Sharp* towed for slightly longer than the F/V *Nordic Pride*. This information was acknowledged and integrated into the modeling portion of the analysis.

Scallop abundance encountered during the experiment varied broadly, however, with resource levels being at historical highs, scallops were encountered at the majority of stations. Any tows that did have a zero catch for both gears were excluded from the analysis as they offered no information relating to the relative efficiency of the two vessels. On a tow-by-tow basis, relative catch varied across the range of stations sampled. This information is displayed in Figures 2-3.

One Parameter Model (Vessel Effect)

The one parameter model (intercept only), utilized the scaled scallop catch that was pooled over all lengths. With this data set, the relative efficiency of the R/V *Hugh Sharp* relative to the F/V *Nordic Pride* with two gear configuration was estimated. Results from this model are shown in Table 3. Estimated relative efficiency values (interpreted as ($\hat{\rho}$ =exp($\hat{\beta}_0$)) were 1.109 and 1.119 for the standard dredge and prototype dredge, respectively. For this model run, $\hat{\beta}_0$ was not statistically significant for either gear configuration. Scatter plots of the scaled scallop catches pooled over all lengths are shown in Figures 4-5. The estimated intercept ($\hat{\beta}_0$) for each vessel/ gear combination is represented as the slope of the dashed line.

Two Parameter Model (Vessel and Length Effects)

The two parameter mixed effect model describes differences in both the number as well as the length of scallops encountered by the two vessels. This model assumes that length varies linearly and allows both the intercept as well as the slope to vary randomly from station to station. This model uses the unscaled data and includes an offset term to account for differential sub-sampling of the catch as well as potential differences in towing distance. With this data set, the relative efficiency of the R/V *Hugh Sharp* relative to the F/V *Nordic Pride* as a function of scallop length was estimated. Model output is shown in Table 4. Plots of observed and predicted proportions overlaid on the scallop length frequency distributions are shown in Figures 6 and 7. Based upon the estimated length based relative efficiency values as calculated parameter estimates plots of the adjusted catch by the R/V *Hugh Sharp* are shown for each gear configuration in Figures 8 and 9. These adjusted catches appear to capture the differences in the raw data given the assumption of the differences in the lengths that vary linearly with respect to length. It is possible that a more complicated relationship (i.e. higher

order polynomials) may be present, although low catches at very small and large scallop lengths may be misleading (Holst and Revill, 2009).

Significant length effects were observed for both gear types tested aboard the F/V *Nordic Pride.* In both of these cases, the length based relative efficiency values increased as a function of increasing scallop size (i.e. positive slope). This positive slope is especially marked for the prototype dredge, where the increasing efficiency as a function of increasing scallop size is observed. The relationship is similar for the standard dredge, however the degree to which the relative efficiency is reduced in a relative sense. Care must be taken to realize that given small sample sizes at both large and small scallop lengths, minimal weight in the likelihood function is given to those observations. Since they are depicted as proportions, the importance of these scallop lengths should be considered with caution. In addition, it is likely that these observations in addition to being at a low sample sizes were likely observed at a small number of stations. This fact reduces their impact on the overall estimation even more.

Discussion

Monitoring the changes in fish and shellfish abundance over time is a critical component of the assessment and management of aquatic resources (Gunderson 1993). Much of the information to accomplish these assessments comes from fishery independent surveys conducted by governmental agencies. As these time series grow older, it becomes more difficult to maintain a standardized survey operation. Vessels age and need to be replaced, and technology improves, necessitating the updating of older gear configurations. While these changes presumably allow the fishery biologist to collect more precise data, care must be taken during times of transition not to introduce a systematic bias into the time series (Pelletier 1998).

In an effort to facilitate the transition of the NMFS NEFSC's sea scallop survey from the retiring F/V *Albatross IV* to the R/V *Hugh Sharp* a calibration experiment was conducted. The commercial sea scallop vessel was selected based on prior work that calibrated this vessel to

the NMFS sea scallop survey time series. A paired design was deemed to be the most expedient approach to calibrate the vessels, and with the standard protocol specifying short tow times (15 min), relatively short steaming times between stations, many tows pairs were accomplished. Wilderbuer *et. al.,* 1998 warns against implementing any correction factor when less than 50 valid tow pairs (non-zero) were used to estimate the correction factor. Based on the operational characteristics of the survey, and the healthy status of the scallop resource, obtaining an adequate sample of valid tow pairs was attained.

Our results indicate a general robustness of the dredge survey to the effect of the vessel. Correction factors were generally small and in all cases, the R/V *Hugh Sharp* was more efficient than the commercial vessel for both gear configurations. These results are generally in agreement with previous work done estimating vessel effect for sea scallop surveys. Throughout the service life of the R/V *Albatross IV*, instances arose where the vessel was not available and a replacement vessel had to be used. In these circumstances, comparative fishing experiments were conducted to justify the use of the surrogate vessel (Serchuk and Wigley, 1989, Lai and Kimura, 2002). These studies found no evidence to support significant differences in relative efficiency between the R/V *Albatross IV* and the replacement vessels used in the comparisons. These studies, however, were somewhat limited in both spatially and temporally.

While a change in the design of a fishing gear would potentially result change in the size composition of the catch, a change in the vessel could also have a similar result. The modeling approach used in this study allowed for an examination of the differences in the length distribution of the catch. Significant differences in the length distributions of the captured scallops were observed between the R/V *Hugh Sharp* and the F/V *Nordic Pride* for both gear configurations. In this study, as well as many traditional calibration analyses, pooled catches (over all lengths) are used to characterize CPUE (Kappenaman, 1992; Lai and Kimura, 2002; Tyson *et. al*., 2006). This pooling of the catch can mask important differences that are vital to

characterize the surveyed population and could potentially result in erroneous assumptions regarding indices of abundance. While attempts have been made to group fish into size classes and test for differences in relative efficiency at this level of resolution, analytical approaches that explicitly model fish length provide stronger statistical inference regarding length based relative efficiency.

Traditional calibration experiments have relied upon standard parametric approaches to analyze the resulting data and estimate FPC factors. The implicit assumption to these analyses is that there is equal availability of animals in the tow path for each vessel and any difference in CPUE is a function vessel effect (Tyson, 2006). Mixed effect models do not necessarily make this assumption and can implicitly model the varying densities encountered by the individual vessels on a tow-by-tow basis (Littell, *et. al.*, 2006). In addition to accounting for the variation which results from differing animal densities, the random effects in the model will absorb another source of variability. Implicit to models that describe fishing gear size selectivity, and catch comparisons, the phenomenon of between haul variation is well described (Fryer, 1991, Millar et al, 2004). This variation is also considered a random effect that can arise from a large variety of sources associated with fishing operations. For this study, the random effects include both between haul variation in the fishing gear as well as the potential differences in scallop densities encountered by the two vessels.

Simulations by Cadigan and Dowden (2009) explored the performance of mixed models for the estimation of relative efficiency from paired tow calibration data. Models that included a random effect performed more reliably in the presence of between tow variation in animal density relative to models that did not include a random effect to account for this source of variability. In this study, we did not evaluate the relative performance of GLMs and GLMMs, but felt that the even though the experimental protocol specified that the vessels in this experiment consecutively occupied the same tow (within the abilities of the vessel operator and constraints of the environmental conditions), satisfying the assumption of equal scallop availability for both

vessels was unrealistic and overly restrictive. In addition, simulations as well as our results indicate that the mixed model was able to accommodate outliers with minimal impact to parameter estimates (Cadigan and Dowden, 2009).

In this experiment we attempted to facilitate the transition of the NMFS sea scallop survey to the current vessel platform. In conjunction with a number of direct comparative tows between the R/V *Albatross IV* and R/V *Hugh Sharp* (the same station locations were sampled on Georges Bank, however, the re-occupations were separated by a number of months), we provided indirect information to calibrate the R/V *Hugh Sharp* in a stepwise fashion. Of primary interest were the catches of the F/*V Nordic Pride* towing the standard dredge. This vessel/gear combination was shown to perform equivalently in a prior paired tow experiment and was assumed to represent the time series. The use of the prototype dredge provided an additional comparison as well as a relative comparison of the performance of the two dredge designs in a side-by-side experiment.

The scope of this experiment covered a broad range of habitats and *in situ* conditions associated with synoptic survey of this species. The modeling approach used in this study utilized a statistical model that was appropriate for count data and the incorporation of a random effect in the model resulted in the ability to account for spatial variation in stock densities. The estimation of FPCs for sea scallops utilizing a commercial vessel with two different survey gears, allows for some latitude going forward. Potentially the commercial vessel could be utilized in future survey work, if the issue of vessel platform again becomes relevant. Regardless, results from both the 2007 study as well as the current work suggest that the scallop survey is fairly robust to vessel effect. While caution should always be exercised when making changes to a major component of the synoptic survey, the expectation of minimal bias stemming from a change in vessel platform is empirically supported. This experiment has provided experimental evidence to support the transition to a new platform while maintaining the

temporal continuity of the time series that has characterized sea scallop abundance since the 1970s.

Literature Cited

Cadigan, N.G., Walsh, S.J and W. Brodie. 2006. Relative efficiency of the *Wilfred Templeman* and *Alfred Needler* research vessels using a Campelen 1800 shrimp trawl in NAFO Subdivisions 3Ps and dividsions3LN. Can Sci Advis Secret Res Doc 2006/085; 59 pp.

Cadigan, N.G. and J. J. Dowden. Statistical inference about relative efficiency from paired-tow survey calibration data. Fish. Bull. 108:15-29.

DuPaul, W.D., E.J. Heist, and J.E. Kirkley, 1989. Comparative analysis of sea scallop escapement/retention and resulting economic impacts. College of William & Mary, Virginia Institute of Marine Science, Gloucester Point, VA. VIMS Marine Resource Report 88-10. 70 pp.

Fryer, R.J. 1991. A model of between haul variation in selectivity. ICES J. of Mar. Sci. **48**: 281-290.

Gunderson, D.R. 1993. Surveys of Fisheries Resources. John Wiley & Sons, Inc. New York, New York.

Helser, T.E., A. Punt and R. Methot. 2004. A generalized mixed model analysis of a multi-vessel fishery resource survey. Fisheries Research. **70**:251-264.

Hilborn, R. and C.J Walters. 1992. Quantitative Fisheries Stock Assessment: Choice Dynamics and Uncertainty. Chapman & Hall, London.

Holst, R. and A. Revill. 2009. A simple statistical method for catch comparison studies. Fisheries Research. **95**: 254-259.

Kappenman, R. F. 1992. Robust estimation of the ratio of scale parameters for positive random variables. Communications in Statistics. Part A: Theory and Methods. **21**:2983–2996.

Kimura, D.K. and D.A. Somerton. 2006. Reviews of statistical aspects of survey sampling for marine fisheries. Reviews in Fishery Science. **14**:245-283.

Lewy, P., J.R. Nielsen and H. Hovgard. 2004. Survey gear calibration independent of spatial fish distribution. Canadian Journal of Fisheries and Aquatic Sciences, **61**: 636-647.

Lia, H.L. and D.K. Kimura. 2002. Analyzing survey experiments having spatial variability with an application to a sea scallop fishing experiment. Fisheries Research. 56: 239-259

Littell, R.C., Milliken, G.A., Stroup, W., Wolfinger, R., and W.O. Schabenberger. 2006. SAS for Mixed Models (2nd ed.). Cary, NC. SAS Institute Inc.

Millar, R.B., M.K. Broadhurst, and W.G. Macbeth. 2004. Modeling between-haul variability in the size selectivity of trawls. Fisheries Research. **67**:171-181.

Munro, P.T. 1998. A decision rule based on the mean squared error for correcting relative fishing power differences in trawl survey data. Fishery Bulletin. **96**:538-546.

Pelletier, D. 1998. Intercalibration of research survey vessels in fisheries: A review and an application. Canadian Journal of Fisheries and Aquatic Sciences. **55**: 2672–2690.

Robson, D. S. 1966. Estimation of the relative fishing power of individual ships. International Commission for the Northwest Atlantic Fisheries, Research Bulletin. **3**:5–15.

Serchuk, F.M. and S.E. Wigley. 1986. Evaluation of USA and Canadian research vessel survey for sea scallops (*Placopecten magellanicus*) on Georges Bank. Journal of Northwest Atlantic Fisheries Science. **7**:1-13.

Sissenwine, M. P., and E.W. Bowman. 1978. An analysis of some factors affecting the catchability of fish by bottom trawls. International Commission for the Northwest Atlantic Fisheries, Research Bulletin. **13**: 81–87.

Tyson, J.T., T.B. Johnson, C.T. Knight, and M.T. Bur. 2006 Intercalibration of research survey vessels on Lake Erie. North American Journal of Fisheries Management. **26:**559-570.

von Szalay, P.G., and E. Brown. 2001. Trawl comparisons of fishing power differences and the applicability to National Marine Fisheries Service and Alaska Department of Fish and Game trawl survey gear. Alaska Fishery Research Bulletin. **8(2)**:85–95.

Wilderbuer, T. K., R. F. Kappenman, and D. R. Gunderson. 1998. Analysis of fishing power correction factor estimates from a trawl comparison experiment. North American Journal of Fisheries Management, **18**: 11–18.

Wolfinger, R. D. 1999. Fitting nonlinear mixed models with the new NLMIXED procedure. *Proceedings of the 24th Annual SAS© Users Group International Conference (SUGI 24).* Paper# 287. 10pp.

The output and code for this paper was generated using the SAS/ STAT software, Version 9.2 of the SAS System for Windows Version 5.1.2600. Copyright © 2002-2008 by SAS Institute Inc.

SAS and all other SAS Institute Inc. products or services are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

Table 1 Characteristics of vessels used in the comparative fishing experiments.

	R/V *Albatross IV*	**F/V *Nordic Pride***
Hailing Port	Lewes, DE	New Bedford, MA
Owner	Univ. of Delaware (UNOLS)	Nordic Fisheries, Inc.
Year Built	2006	1987
LOA (ft.)	146	92.7
Hull Depth (ft.)	9.5	13.2
Hull Breadth (ft.)	32	26
Gross Tonnage	495	192

Table 2 Environmental and catch information for the two comparative cruises. Scallop catch represents the scaled catch for the valid tows sampled during the two legs of the experiment.

Experimental Vessel	**F/V *Nordic Pride***	
Sampling Location	Mid-Atlantic Bight	
Dates of Cruise	May 9, 2009 to May 17, 2009	
Depth		
Minimum	23	
Maximum	39	
Wind velocity (kts.)		
Minimum	0	
Maximum	20	
Wave height (ft.)		
Minimum	0	
Maximum	5	
Vessel Speed (kts.)	3.8	
Scope	3:1	
Experimental gear	Standard Dredge	Prototype Dredge
Comparative tows	100	100
Total catch (F/V)	60,038	60,200
Total catch (R/V)	69,491	69,491

Table 3 Mixed effects model (vessel effect only) results including an offset term to account for the effect of differential tow lengths. Parameter estimates are on the logit scale and significant estimates are shown in bold.

Vessel/Gear	σ^2	Estimate (β_0)	Standard Error	Lower 95% CI	Upper 95% CI	*t*	*p*-value
R/V *Hugh Sharp* vs. F/V *Nordic Pride* (Standard Dredge)	0.4827	0.1040	0.0707	-0.0364	0.2444	1.47	0.1448
R/V *Hugh Sharp* vs. F/V *Nordic Pride* (Prototype Dredge)	0.5675	0.1131	0.0766	-0.0389	0.2651	1.48	0.1430

Table 4 Two parameter mixed effects model results. Both comparisons model the logit of the proportion of the catch at length from the R/V relative to the total catch from both vessels. Parameter estimates reflect a model that includes an offset term in the model that accounted for both sub-sampling of the catch as well as differences in within-tow areal coverage. Confidence limits are Wald type confidence intervals. Parameter estimates are on the logit scale and significant parameter estimates are shown in bold.

Vessel	DF	σ^2 (intercept)	σ^2 (slope)		Estimate	Standard Error	Lower 95% CI	Upper 95% CI	*t*	*p*-value
R/V *Hugh Sharp* vs. F/V *Nordic Pride* (Standard Dredge)	98	0.4887	0.3802	β_0	0.0908	0.07157	-0.05188	0.2329	1.27	0.2073
				β_1	**0.1184**	**0.06879**	**0.05187**	**0.3249**	**2.74**	**0.0073**
R/V *Hugh Sharp* vs. F/V *Nordic Pride* (Prototype Dredge)	98	0.5980	0.5676	β_0	0.1154	0.07897	-0.04129	0.2721	1.46	0.148
				β_1	**0.3923**	**0.08166**	**0.2302**	**0.5543**	**4.80**	**<0.0001**

Figure 1 Station locations for the mid-Atlantic portion of the 2009 NEFSC sea scallop survey conducted by the R/V *Hugh Sharp*. Comparative stations for the calibration study were a subset of the entire survey and were intended to cover the spatial extent of the domain with its varied biotic and abiotic conditions. Polygons in both areas represent closed areas in existence at the time of the study, which are part of the spatial management strategy for the fishery. Map courtesy of the NEFSC Ecosystem Surveys Branch.

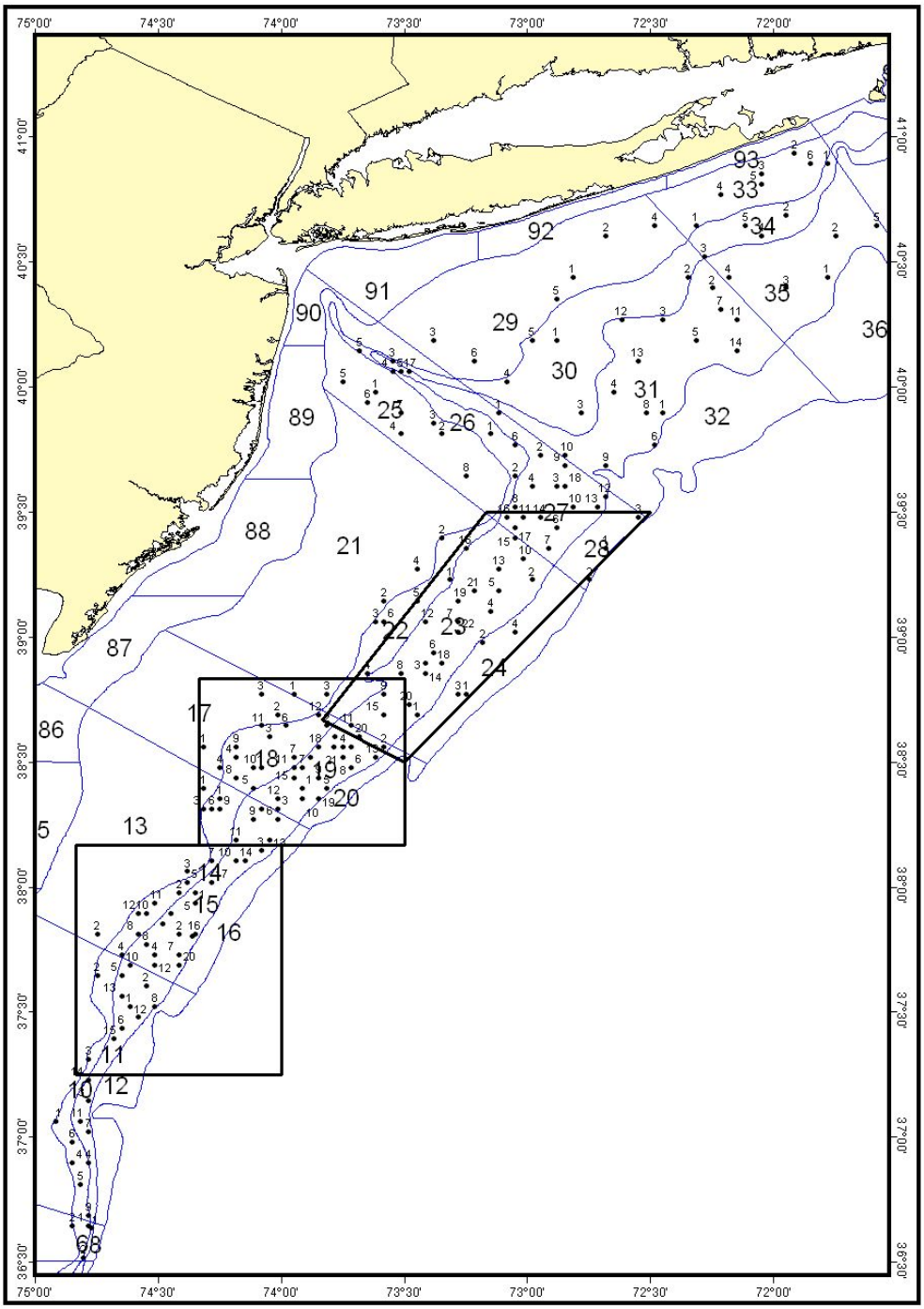

Figure 2 Catch differences between the F/V *Nordic Pride* (towing the standard NMFS dredge) and the R/V *Hugh Sharp*. Catches for each vessel are scaled to reflect both any sub-sampling of the catch as well as differences in areal coverage. Symbols are proportional to the magnitude of the observed differences in catch. Red dots represent higher levels of catch by the R/V *Hugh Sharp*. Blue dots represent higher levels of catch by the F/V *Nordic Pride*. Open circles represent zero difference between the two vessels. Polygons in both areas represent closed areas in existence at the time of the study, which are part of the spatial management strategy for the fishery. The dotted line represents the 50 fathom bathymetric contour.

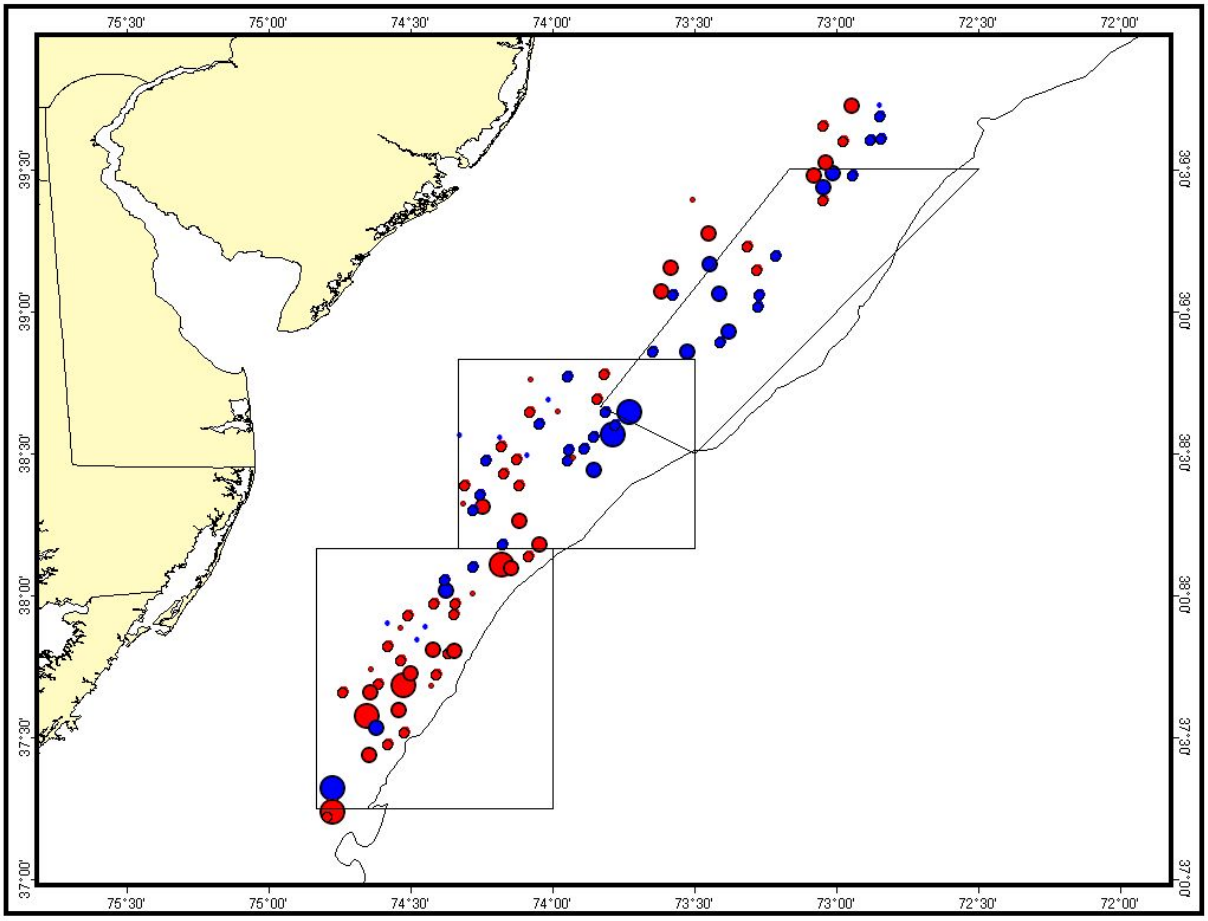

Figure 3 Catch differences between the F/V *Nordic Pride* (towing the prototype NMFS dredge) and the R/V *Hugh Sharp*. Catches for each vessel are scaled to reflect both any sub-sampling of the catch as well as differences in areal coverage. Symbols are proportional to the magnitude of the observed differences in catch. Red dots represent higher levels of catch by the R/V *Hugh Sharp*. Blue dots represent higher levels of catch by the F/V *Nordic Pride*. Open circles represent zero difference between the two vessels. Polygons in both areas represent closed areas in existence at the time of the study, which are part of the spatial management strategy for the fishery. The dotted line represents the 50 fathom bathymetric contour.

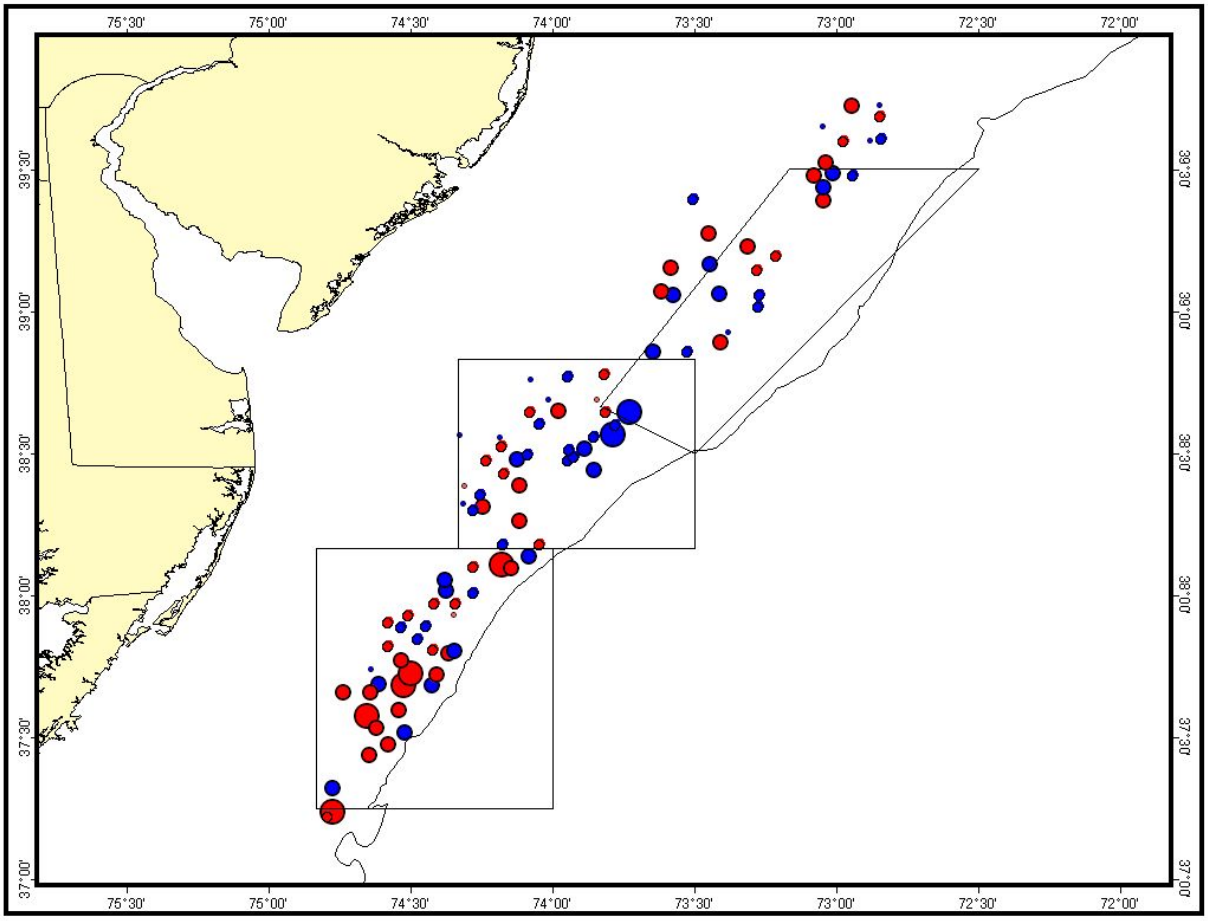

Figure 4 Total scaled catches for R/V *Hugh Sharp* vs. F/V *Nordic Pride* (towing the NMFS standard dredge). The red line has a slope of one. The dashed line has a slope equal to the estimated relative efficiency (from the one parameter vessel effect only model).

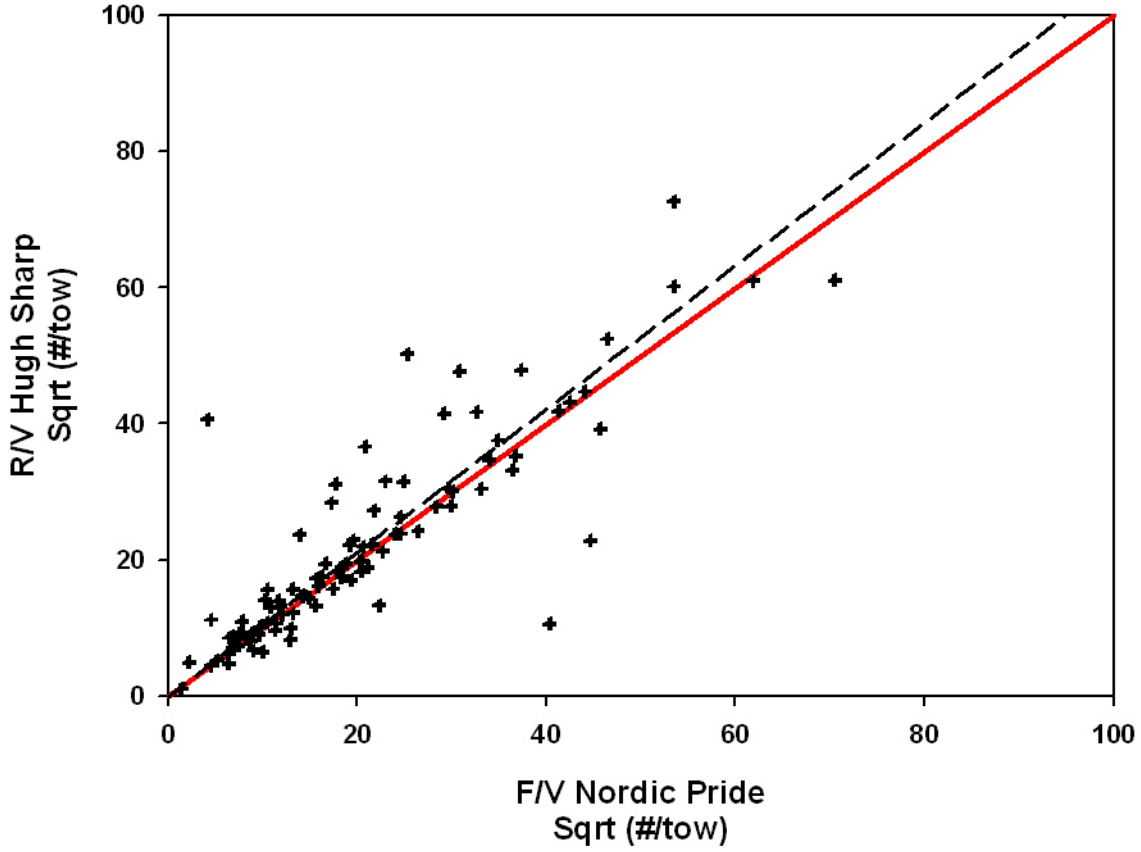

Figure 5 Total scaled catches for R/V *Hugh Sharp* vs. F/V *Nordic Pride* (towing the NMFS prototype dredge). The red line has a slope of one. The dashed line has a slope equal to the estimated relative efficiency (from the one parameter vessel effect only model).

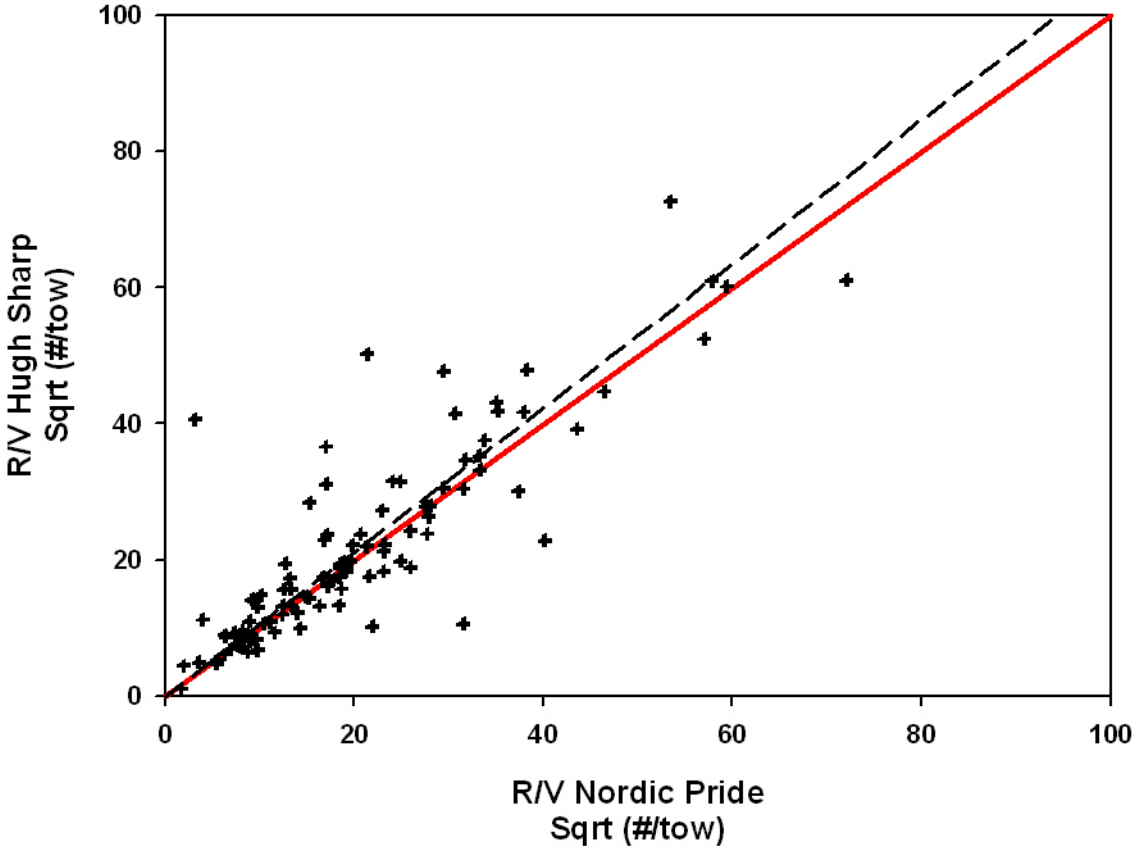

Figure 6 Observed scaled length frequency distributions for the R/V *Hugh Sharp* and the F/V *Nordic Pride* (towing the NMFS standard dredge). The green triangles represent the observed proportions ($Catch_{R/V}$/($Catch_{R/V}$ + $Catch_{F/V}$). The black line represents the length based relative efficiency as estimated by the two parameter (vessel and length effect model).

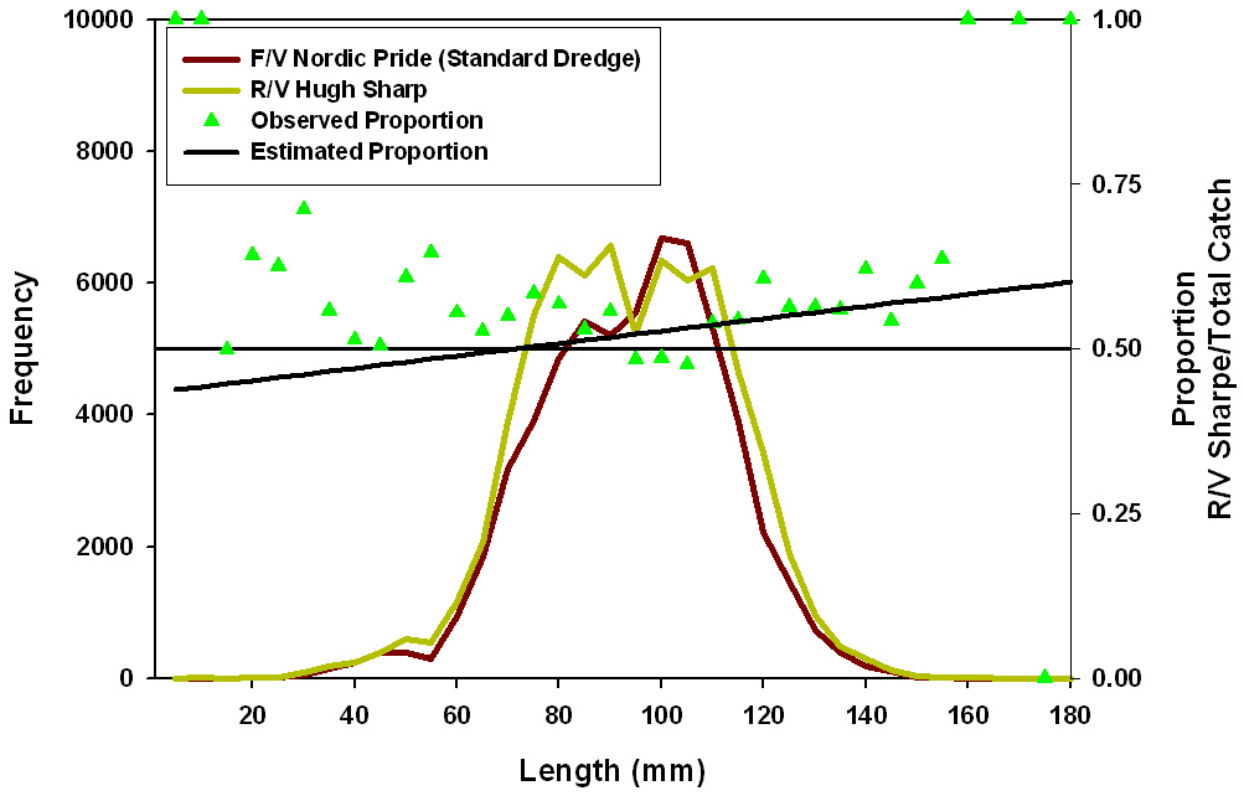

Figure 7 Observed scaled length frequency distributions for the R/V *Hugh Sharp* and the F/V *Nordic Pride* (towing the NMFS prototype dredge). The green triangles represent the observed proportions ($Catch_{R/V}$/($Catch_{R/V}$ + $Catch_{F/V}$). The black line represents the length based relative efficiency as estimated by the two parameter (vessel and length effect model).

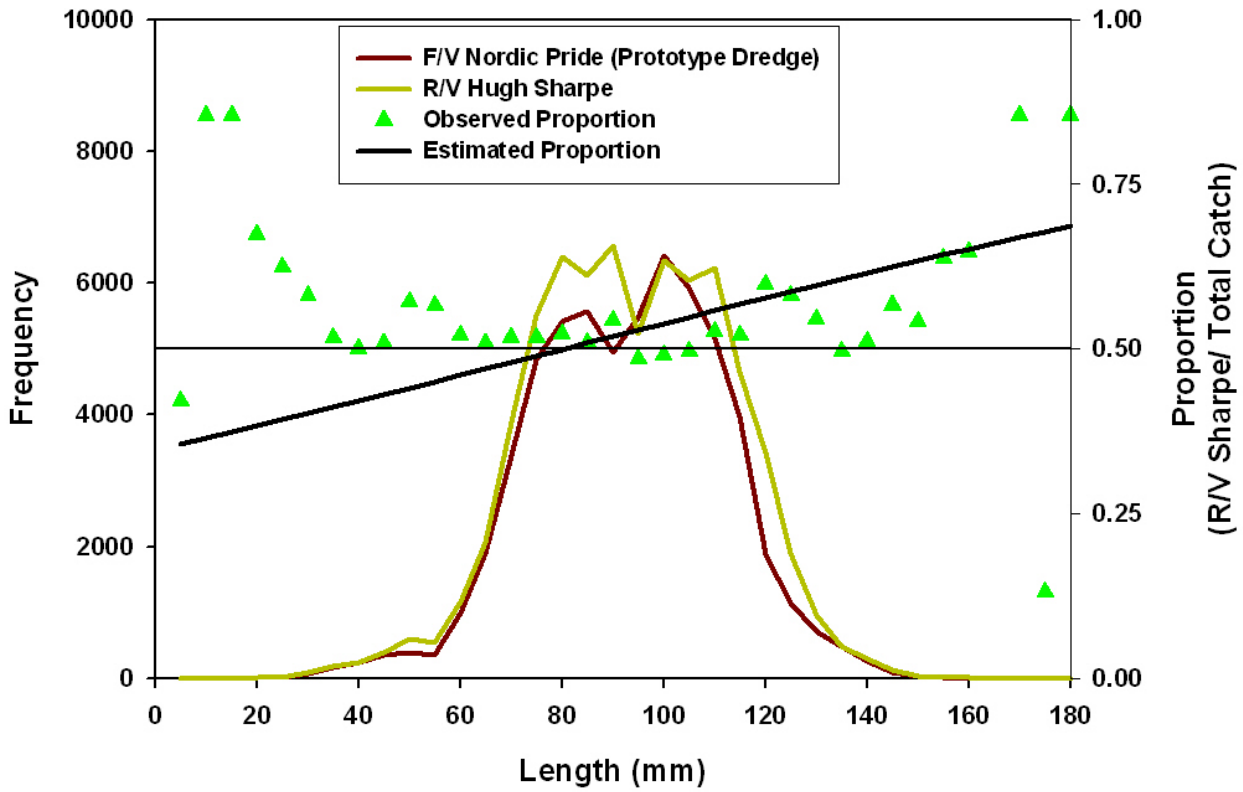

Figure 8 Observed scaled length frequency distributions for the R/V *Hugh Sharp* and the F/V *Nordic Pride* (towing the NMFS standard dredge). The black dashed line represents the adjusted R/V *Hugh Sharp* catches. The adjusted catches were calculated from the length based relative efficiency values calculated by the two parameter (vessel and length effect model).

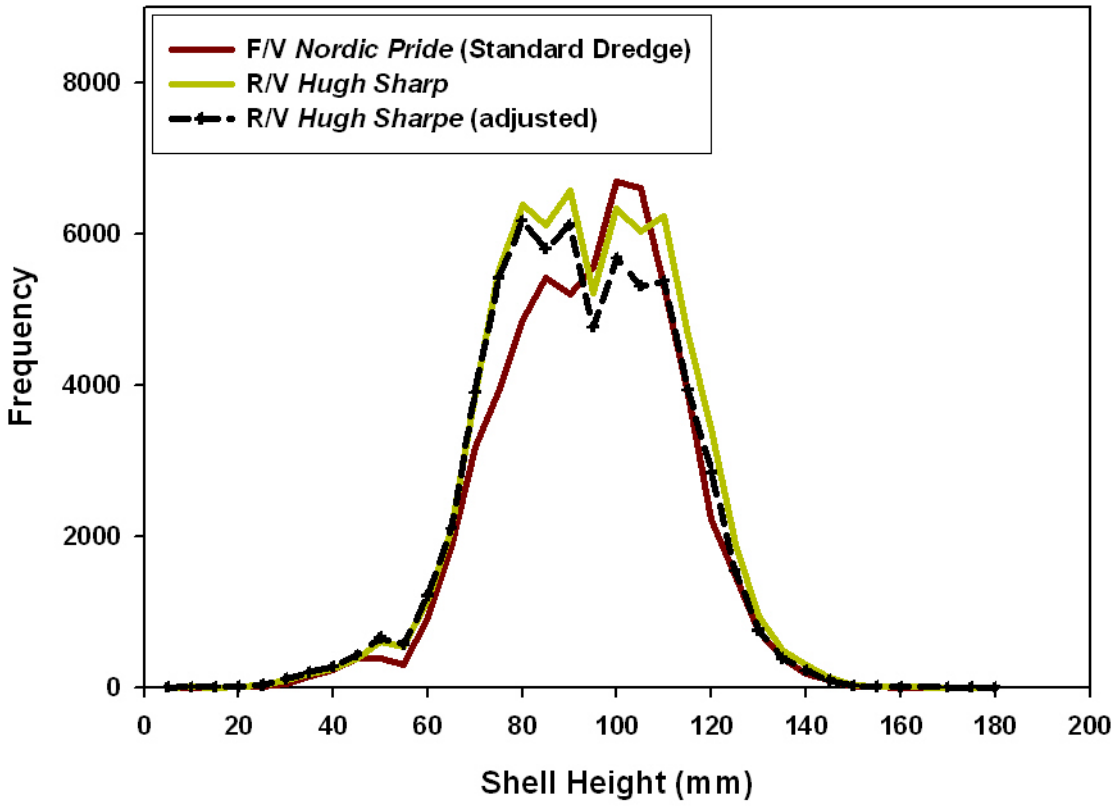
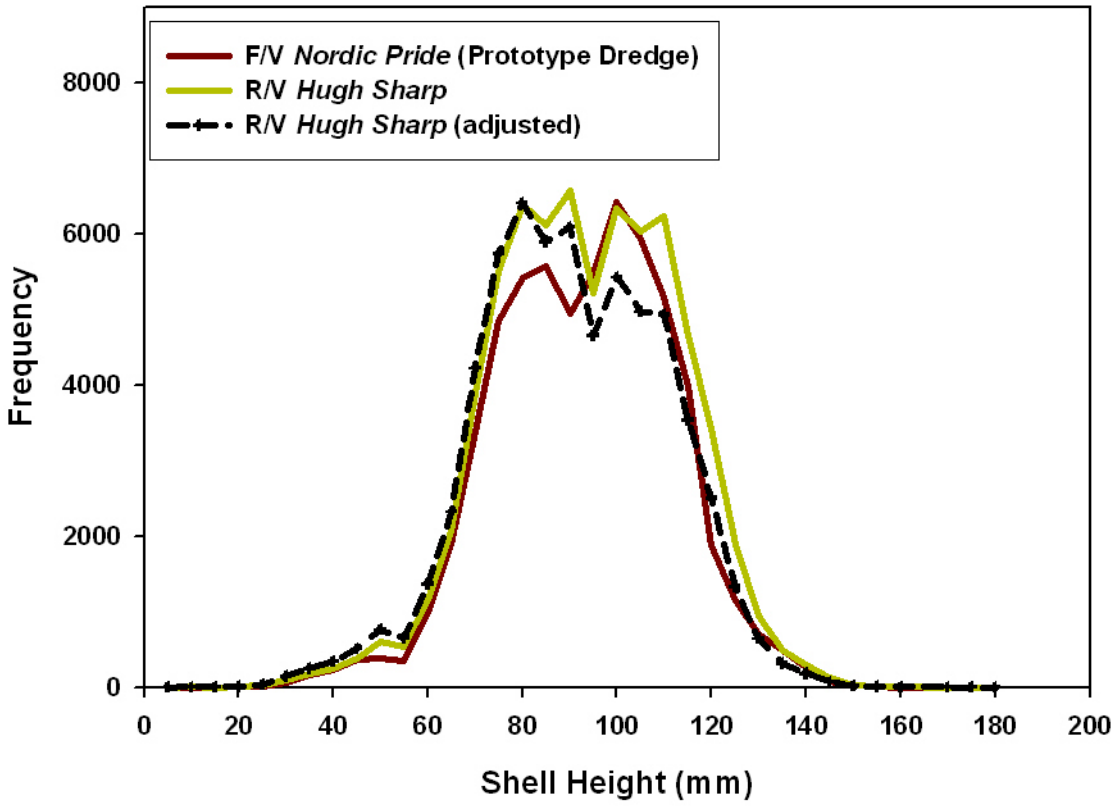
Figure 9 Observed scaled length frequency distributions for the R/V *Hugh Sharp* and the F/V *Nordic Pride* (towing the NMFS prototype dredge). The black dashed line represents the adjusted R/V *Hugh Sharp* catches. The adjusted catches were calculated from the length based relative efficiency values calculated by the two parameter (vessel and length effect model).