Appendix B4: Vessel calibrations for the NMFS sea scallop survey

In anticipation of the retirement of the R/V *Albatross IV*, the NOAA vessel that had conducted the annual synoptic sea scallop survey virtually uninterrupted since the 1970's, a series of paired tow calibration experiments were conducted to estimate fishing power correction factors. The objective of these experiments was to facilitate the transition of the NMFS sea scallop dredge survey time series from the R/V *Albatross IV* to a future survey platform. Due to some uncertainty in the subsequent survey platform, this information would facilitate the use of the calibrated vessel to either conduct the survey, or at least form a link from the R/V *Albatross IV* to any future survey platform. Ultimately, two calibration experiments were conducted in 2007 and 2009 with the calibration process being conducted in a stepwise fashion. We used a Generalized Linear Mixed Model (GLMM) to analyze the paired catch data to test for differences in both the pooled over length catch data as well as differences in the length composition of the catch. In 2007, the commercial scallop vessel, F/V *Nordic Pride* conducted a paired tow experiment with the R/V *Albatross IV*. Results indicate that while the R/V *Albatross IV* was slightly more efficient, the difference was small (~5%) and not statistically significant. Based on these results, the F/V *Nordic Pride* was considered to be equivalent with respect to fishing power to the R/V *Albatross IV*. In 2008, the R/V *Hugh Sharp* was selected as the replacement vessel for the R/V *Albatross IV* and during the 2009 survey an additional paired tow experiment was conducted between this vessel and the F/V *Nordic Pride*. Results indicate that the R/V *Hugh Sharp* was slightly more efficient (~10%) than the F/V *Nordic Pride*, however, this difference was not statistically significant. These results indicate that scallop dredge catches are robust to the effect of vessel and that any correction factor applied to this time series moving forward is small (~5%) or not justified.

Data collection and analysis

Experimental Design

The calibration experiments were conducted within the context of the NMFS annual sea scallop survey. This survey utilizes a stratified random design to sample throughout the entire U.S. range of the sea scallop. (Serchuk and Wigley 1986). For both paired tow experiments, the sampling occurred during the mid-Atlantic portion of the NMFS survey. For the first experiment, the standard NMFS sea scallop survey dredge that has been in service, virtually unmodified since the 1970's was used aboard both vessels. This dredge is 8 ft in width, with a dredge bag consisting of 2 inch rings. The twine top is comprised of 3.5 inch diamond mesh and there is a 1.5" liner throughout the dredge bag. For the second experiment, the F/V *Nordic Pride* used the standard dredge, while the R/V *Hugh Sharp* used a slightly modified version of the standard dredge 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, 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. In essence, the fishing power correction factor estimated for the second experiment attempts to calibrate the existing time series to a new entity that is represented by a unique vessel/gear combination.

While at sea, the sampling protocol included the re-occupation of sampling stations occupied by the R/V *Albatross IV*. Start/stop locations for each tow completed by the R/V *Albatross IV* 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 *Albatross IV* conducts 24 hour operations, while the F/V's in this study sampled for roughly 16-18 hrs/day). During the execution of the tow, the captain of the F/V 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.

For each paired tow, the entire scallop catch was placed in baskets. A fraction of these baskets will be measured to estimate length frequency for the entire catch. The shell height of each scallop in the sampled fraction will be 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 was quantified, with finfish being sorted by species and measured to the nearest 1 cm. Sampling protocol was similar on the R/V *Albatross IV*.

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 and is based on the analytical approach included in Cadigan *et al.,*2006. Assume that each vessel/gear combination tested in this experiment has a unique catchability. Let q_r equal the catchability of the R/V 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 *Albatross IV* or 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 i^{th} station by the R/V and λ_{if} the standardized scallop density encountered by the F/V. 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 capture is a Poisson process with mean equal to variance, then the expected catch by the commercial vessel is given by:

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

The catch by the R/V *Albatross IV* is also a Poisson random variable with:

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

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

If the vessels encounter the same scallop density for a given tow, (i.e. λ_{ir}= λ_{if}), then ρ can be estimated via a Poisson generalized linear model (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:

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

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_ip and $Var(C_i r)$=c_ip/(1-p). Therefore:

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

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:

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

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(β_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:

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

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

In some cases, we encountered difficulties with model convergence for the two parameter model. To simplify the computations in the optimization routine, scallop lengths were standardized to sum 0 based on the interquartile range. This reduced the magnitude of the steps between successive lengths and alleviated the convergence issues. We used SAS/STAT® PROC NLMIXED to fit the generalized linear mixed effects models.

Results and Discussion

Overall, roughly 100 paired tows were completed for each experiment. A visual representation of the spatial distribution of the relative catces for both experiments is shown in Figure 1. For the intercept only model (vessel effect only) a scatterplot of the catches from the paired tows are shown in Figure 2 and parameter estimates are shown in Table 1. For each experiment the R/V was slightly more efficient than the F/V *Nordic Pride* (correction factor is interpreted as $\exp(B_0)$). The calculated correction factors were 1.058 and 1.110 for the two experiments, respectively. In both cases, the logit of the estimated intercept was not significantly different than 0.

For the two parameter model (length effects) there was a significant difference detected in the length composition of catches from the two vessels (Figure 3 and Table 2). The direction of the difference was consistent between the two experiments and showed that the R/V was more efficient as a function of increasing scallop length. The increase in relative efficiency with respect to length for the first cruise may have resulted from measurement errors associated with different measuring devices between the two vessels. For the second experiment, an apparent pattern in the residuals at the small lengths was apparent, however the sum of the animals from lengths <60 mm only represented roughly 4% of the total catch and likely contributed little weight in the likelihood.

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. This work is not yet in press, but is expected to be at any time)

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

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, W.G. Macbeth. 2004. Modeling between-haul variability in the size selectivity of trawls. Fisheries Research. **67**:171-181.

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.

Table 1 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
F/V Nordic Pride vs. R/V *Albatross IV*	0.2386	0.0568	0.0501	-0.0427	0.1562	1.13	0.2602
F/V Nordic Pride vs. R/V *Hugh Sharp*	0.4827	0.1040	0.0707	-0.0364	0.2444	1.47	0.1448

Table 2 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
F/V Nordic Pride vs. R/V *Albatross IV*	98	0.2744	0.5077	β_0	0.01199	0.05454	-0.09625	0.1202	0.22	0.8264
				β_1	**0.4983**	**0.07964**	**0.3402**	**0.6563**	**6.26**	**<0.0001**
F/V Nordic Pride vs. R/V *Hugh Sharp*	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**

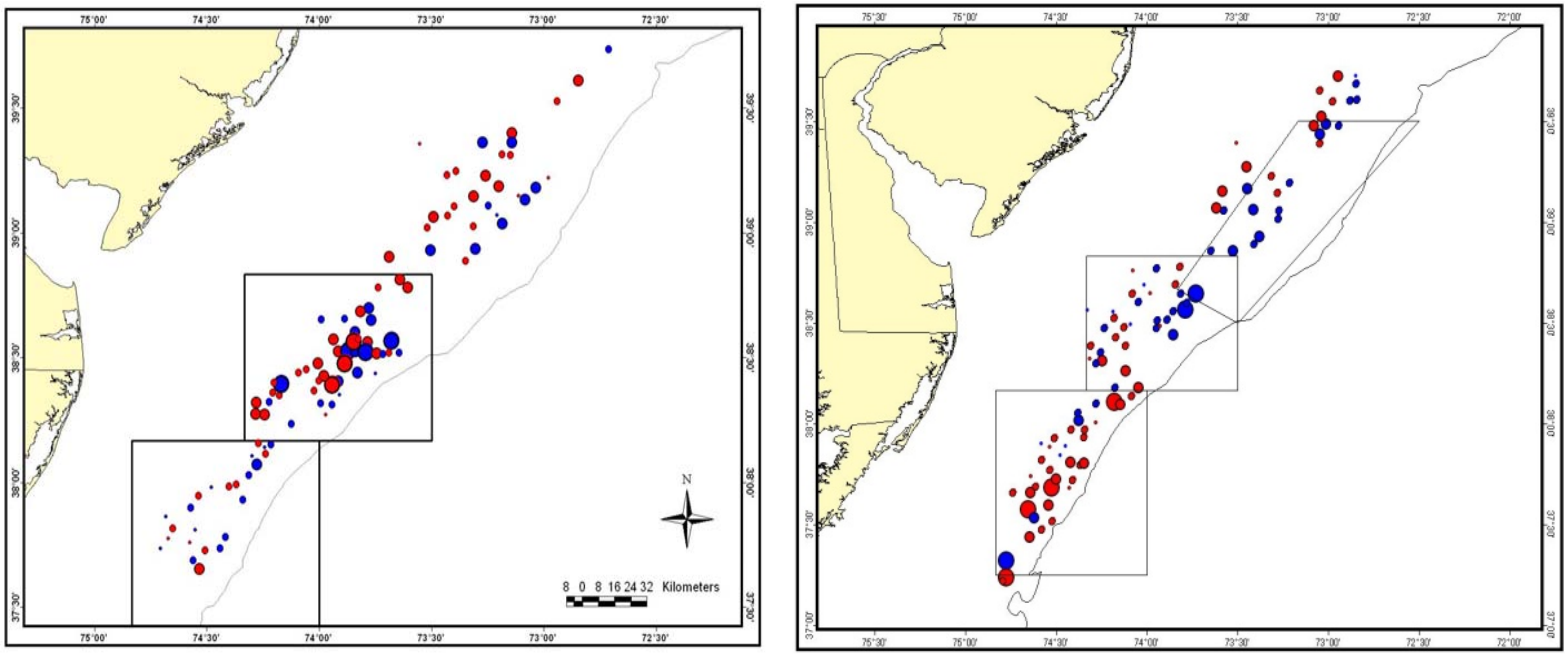
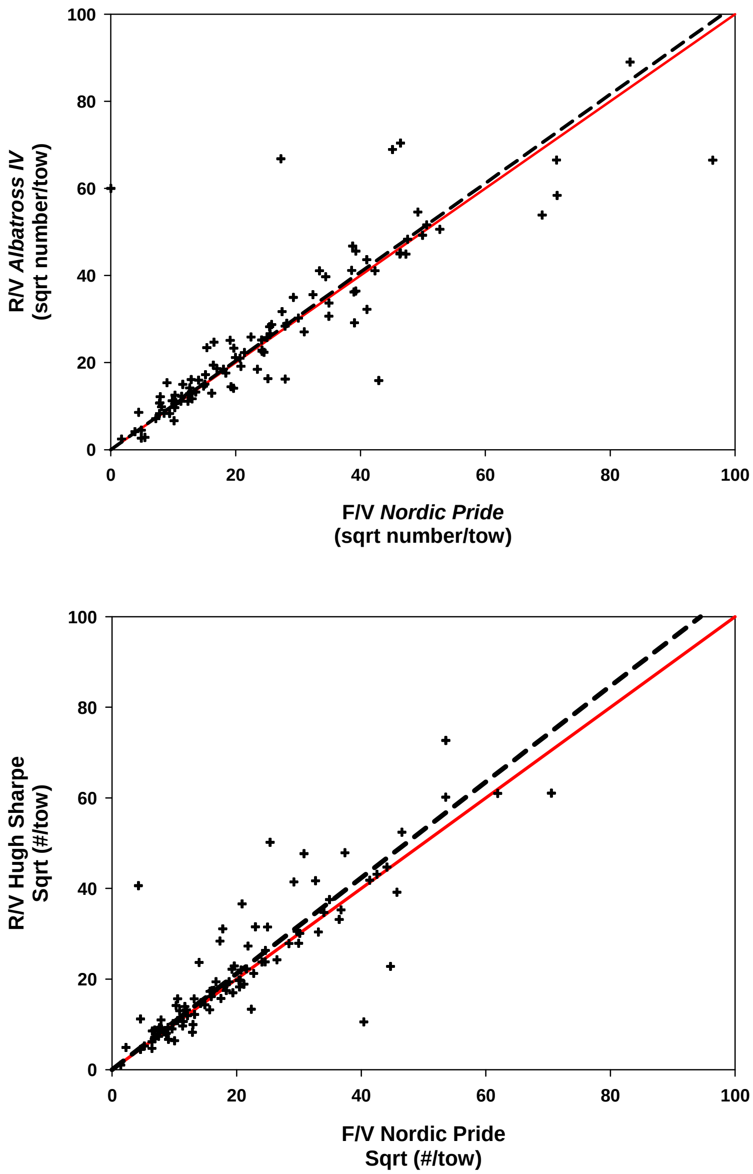


Figure 1 Catch differences between the F/V *Nordic Pride* (towing the standard NMFS dredge) and the R/V *Albatross IV* (left panel) or the R/V *Hugh Sharp* (right panel). Catches for each vessel are scaled to reflect both any sub-sampling of the catch as well as differences in areal coverage. Symbols are proportional to the magnitude of the observed differences in catch. Red dots represent higher levels of catch by the R/V. 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.



<u>Figure 2</u> <u>Top Panel:</u> Total scaled catches for R/V *Albatross IV* vs. F/V *Nordic Pride* (top panel) and the R/V *Hugh Sharp* vs. the F/V *Nordic Pride* (bottom panel). 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).

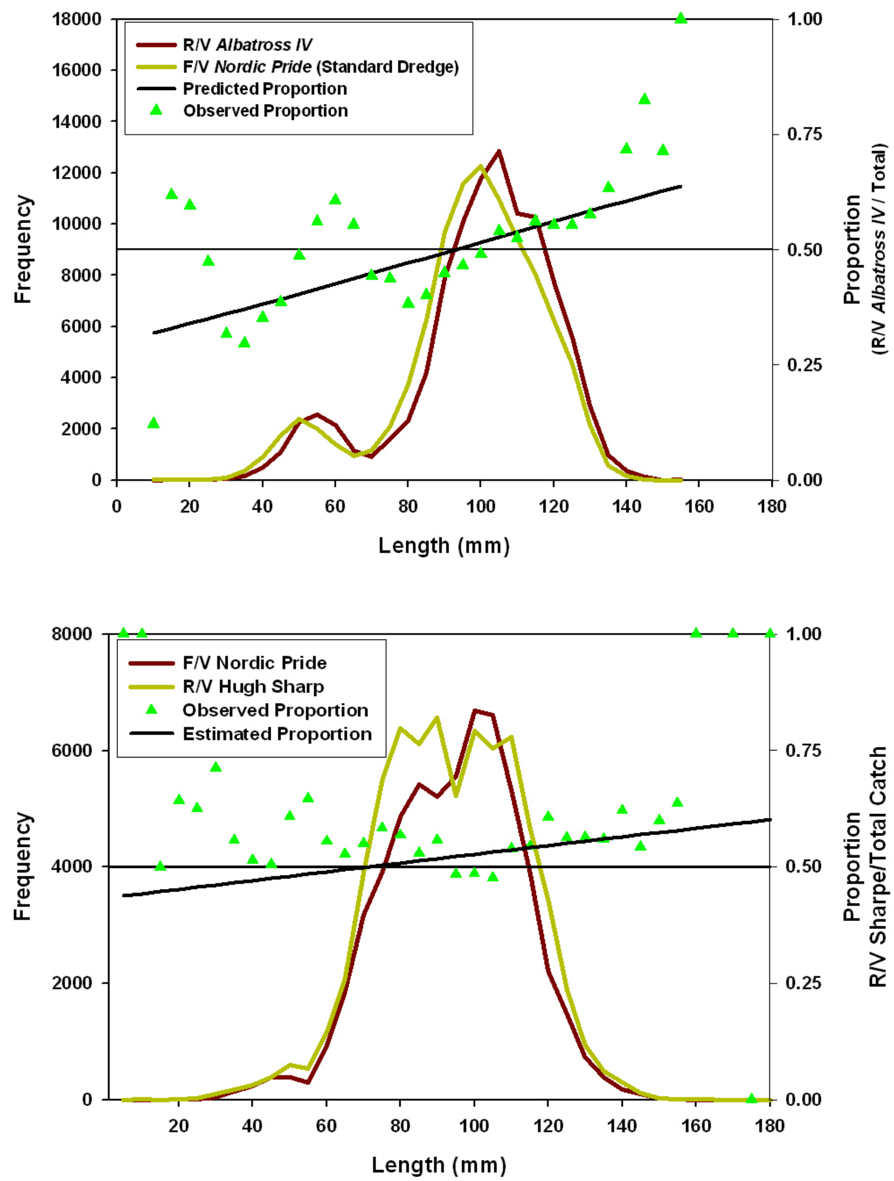


Figure 3 Observed scaled length frequency distributions for the R/V *Albatross IV* and the F/V *Nordic Pride* (top panel) and the R/V *Hugh Sharp* and F/V *Nordic Pride* (bottom panel). The green triangles represent the observed proportions ($Catch_{R/V}/(Catch_{F/V} + Catch_{F/V})$). The black line represents the length based relative efficiency as estimated by the two parameter (vessel and length effect model.