Fisheries and Oceans Canada / Pêches et Océans Canada
Science / Sciences

C S A S — Canadian Science Advisory Secretariat

S C C S — Secrétariat canadien de consultation scientifique

Research Document 2006/085

Document de recherche 2006/085

Not to be cited without permission of the authors *

Ne pas citer sans autorisation des auteurs *

Relative efficiency of the *Wilfred Templeman* and *Alfred Needler* research vessels using a Campelen 1800 shrimp trawl in NAFO Subdivision 3Ps and Divisions 3LN

Étude de l'efficacité relative des navires Wilfred Templeman et Alfred Needler utilisant le chalut à crevettes 1800 Campelen dans la sous-division 3Ps et les divisions 3LN de l'OPANO

N.G. Cadigan , S.J. Walsh, and W. Brodie

Science Branch
Department of Fisheries and Oceans
P.O. Box 5667, St. John's, NL
Canada. A1C 5X1

ISSN 1499-3848 (Printed / Imprimé)

Canada

ABSTRACT

In many of the multi-species trawl surveys conducted by the Newfoundland Region of Fisheries and Oceans Canada, the survey vessel "Wilfred Templeman" (WT) may be replaced by the vessel "Alfred Needler" (AN). We examined paired-trawl experiments involving these two vessels to examine for differences in catchability when both vessels fish the Campelen 1800 shrimp trawl which is the standard survey trawl used in the Newfoundland Region. The location for the comparative fishing was in NAFO Subdivision 3Ps and Divisions 3LN. The results overall suggested that there were no significant differences in the relative catchability of the two vessels. We also presented a generalized linear mixed effects model with an auto-correlated random effect that we suggest is useful for estimating relative efficiency, or the ratio of catchability, from paired-trawl survey calibration data when there is substantial local variability in stock abundance fished by each vessel. We compared these estimates with those from a more commonly used approach involving standard logistic regression. We found the mixed model approach fit the data better and produced estimates of relative efficiency that were not heavily influenced by a small number of outliers. The mixed model results indicated that differences in catchability were small and not statistically significant. However, the logistic regression approach produced some small but statistically significant estimates, and the significance seemed suspect because the effects were not apparent in graphical analyses.

RÉSUMÉ

Dans bon nombre de relevés au chalut plurispécifiques menés par Pêches et Océans Canada dans la région de Terre-Neuve, le navire Wilfred Templeman (WT) pourait être remplacé par le navire Alfred Needler (AN). Nous avons examiné les expériences de chalutage jumelées auxquelles ont participé ces deux navires pour étudier leurs différences de capturabilité lorsqu'ils utilisent tous les deux le chalut à crevettes Campelen 1800, lequel est normalement utilisé dans la région de Terre-Neuve. La pêche comparative a eu lieu dans la sous-division 3Ps et les divisions 3LN de l'OPANO. Les résultats montrent que, dans l'ensemble, il n'y a aucune différence importante entre la capturabilité relative des deux navires. Nous présentons également un modèle linéaire à effets mixtes généralisé avec effet aléatoire autocorrélé que nous estimons être utile pour l'estimation de l'efficacité relative, ou le rapport de capturabilité, au moyen de données calibrées d'après des relevés au chalut jumelés lorsqu'il y a une variabilité locale importante de l'abondance des stocks prélevés par chaque navire. Nous avons comparé ces estimations à celles obtenues à l'aide de l'approche de régression logistique standard couramment utilisée et constaté que le modèle à effets mixtes s'ajustait mieux aux données et produisait des estimations de l'efficacité relative non fortement influencées par un faible nombre de valeurs aberrantes. Les résultats du modèle à effets mixtes montrent que les différences dans la capturabilité sont faibles et non statistiquement significatives. Cependant, l'approche de régression logistique a produit des estimations de faible valeur mais statistiquement significatives, quoique leur signification paraisse douteuse, les effets n'étant pas apparents dans les analyses graphiques.

INTRODUCTION

A fish stock assessment involves evaluating the current status of a stock relative to its past, and evaluating the past consequences of commercial fishing on the stock. Stock assessments may also provide projections or forecasts of future stock status and the consequences of future fishing on the stock. Stock indices are fundamental components of stock assessments. An index is a "measurement" that we usually expect to be proportional to stock size.

A random index R_y available for year y is related to stock size (S_y) via the model

$$E(R_y) = qS_y. \tag{1}$$

We treat S_y as a fixed quantity to estimate. The constant of proportionality, q, is usually referred to as the index catchability, and q should be the same from year to year. Although we cannot directly infer stock size from a time series of indices $R_1, ..., R_Y$, we can infer trends in stock size when q is the same each year. Note that q may be much different from one for many reasons; for example, the index may be based on a fishing gear that does not catch small fish, or the index may be based on measurements from only part of the stock area.

Stock size indices are often based on a survey in which randomly chosen sites are fished. We focus on stratified random bottom trawl surveys such as those conducted off the east coast of Canada by Fisheries and Oceans Canada (e.g. Doubleday 1981). These are multi-species surveys that are used extensively in stock assessments. The information collected from these surveys is used for many other purposes as well, such as determining species at risk (e.g. Smedbol et. al. 2002) and evaluations related to closed areas. The survey observation is commonly referred to as a *set* (i.e. *set* the gear), or a *tow* when a trawl is used. The average survey catch can be taken as an index of stock size. If the same survey protocols are used from year to year then the catchability of the index should remain relatively constant.

In this paper we examine if a change in survey vessels has an impact on stock size indices derived from survey bottom trawling. In many of the multi-species trawl surveys conducted by Fisheries and Oceans Canada, the survey vessel "*Wilfred Templeman*" (WT) may be replaced by the vessel "*Alfred Needler*" (AN) . Although these "sister" vessels are similar in construction, the AN has a larger engine size (2600 horsepower) compared to the WT (2000 horsepower). Hence, there is a potential that they have different catchabilities. Vessel differences in survey catchabilities, even when using the same gear, are common (e.g. Cotter 2001, Pelletier 1998, Wilderbuer and Kappenman 1998). Vessel differences in trawl geometry and swept area using the same fishing gear were observed by McCallum and Walsh (2002). We examined paired-trawl experiments to estimate the relative difference between WT and AN catchabilities, when both vessels fished the Campelen 1800 shrimp trawl, which is the standard survey trawl used by the Newfoundland Region of Fisheries and Oceans Canada.

In paired-trawl experiments two vessels are fished as close together as possible to minimize spatial heterogeneity between the stock densities the vessels encounter; therefore, differences in catches should primarily reflect differences in catchabilities. Pelletier (1998) reviewed estimation methods used in many vessel calibration experiments. In the past a common approach was to log transform catches and use normal linear models for analysis; however, this approach does not often adequately account for the stochastic nature of the data (e.g. counts) and involves arbitrary choices to deal with zero catches. A better approach is to treat the catches from both vessels as Poisson or over-dispersed Poisson random variables, which are statistical distributions that are appropriate for count data, including zero counts. This approach was used by Benoît and Swain (2003), although it is complex because many fish density parameters for each tow usually have to be estimated. Pelletier (1998) used a similar approach, with a mean-variance assumption that is the same as an over-dispersed Poisson distribution (i.e. Negative Binomial). Pelletier (1998) suggested that the number of parameters to estimate may be reduced in some situations by assuming fish densities are constant between paired tows, or that the densities are random with the same mean between tows. However, this assumption will not be appropriate in a large-scale paired-trawl comparative fishing experiment.

It is much simpler to use an associated conditional distribution that treats the total catch-at-length from both vessels as fixed. This eliminates the large number of fish density parameters, and the corresponding statistical likelihood function only involves relative catchability parameters. We describe this approach in more detail in the Methods. The sums of catches-at-length from both trawls are treated like sample sizes. This conditional approach is commonly used in commercial fishing gear size-selectivity studies (e.g. Millar 1992) and has been used in paired-trawl calibration studies (e.g. Fanning 1985, Lewy et. al. 2004). Paired-trawl size-selectivity experiments are essentially the same as the type of experiment we consider.

We also address local spatial variability in stock densities fished by each trawler, or within-pair variability. Although vessels are fished close together in a paired-trawl experiment, it is not possible to ensure that exactly the same stock densities are fished by both vessels. This produces within-pair residual correlation in trawl catches. Some of the correlation could also be due to between-set variability in relative efficiency. It is well-known in gear size-selectivity studies that a failure to account for between-set variability leads to confidence intervals for selectivity parameters that are too narrow, and spurious statistical significance is likely to be observed (Millar et. al. 2004, Fryer 1991). Similar problems have been reported for survey calibrations studies (e.g. Benoît and Swain 2003). We will not be able to separate variability due to within-pair differences in stock densities and variability due to between-pair differences in relative efficiency so we simply assume for convenience that the extra variability is due to stock densities. Our results should be valid in either case (see Discussion).

Benoît and Swain (2003) and Lewy et. al. (2004) used an over-dispersion parameter to account for extra-Poisson or extra-Binomial variation, although they were not clear about what the source of the extra variation was. Cotter (2001) noted the cluster sample nature of trawl survey data and adjusted a variance estimate to account for the cluster sampling. We do not use his approach to estimate relative efficiency for reasons outlined in the Discussion. Note that Lewy et. al. (2004) advocated paired-trawls along the same trawl track line to avoid complications due to spatial variations in stock densities. However, such trawling introduces a different complication, which involves disturbance of the fish densities encountered by the second vessel because of the fishing activity of the first vessel.

We account for cluster sampling and local spatial variability in stock densities fished by each trawler using a generalized linear mixed model (GLMM), which is a class of hierarchical models that provide a flexible parametric approach for the estimation of covariate effects with clustered data. A general description is given in, for example, McCulloch and Searle (2001). GLMM's are particularly useful for structuring multiple sources of variation, both measured (covariates) and unmeasured (random effects). The use of a nonlinear link function in a GLMM means that proper selection of the random effects structure may be required for valid point estimates and for correct standard errors. A simple over-dispersion parameter is often not sufficient to account for many types of random effects. Heagerty and Kurland (2001) demonstrated that large biases in regression parameter estimates can occur when random effects are misspecified. We consider their results further in the Discussion; however, Heagerty and Kurland (2001) recommend that "careful attention be given to the random effects model assumptions when using generalized linear mixed models for regression inference with clustered or longitudinal categorical data". In the next section we do this for paired-trawl calibration studies.

METHODS

The main objective of the comparative fishing exercise was to determine if differences exist between WT and AN catches when both vessels used the standard Campelen 1800 survey trawl. Data from paired tows were collected to quantify potential differences. Ranges of catch sizes, fish sizes in the catch, and tow depths were sought, given the distributions of the species likely to be encountered. The location of the comparative fishing was in NAFO Subdiv. 3Ps and Div. 3LN. Tow stations were selected randomly as part of research surveys. High density aggregations were not specifically targeted because information was required on differences in catchability when stock densities are both high and low, which typically occurs in research surveys.

PAIRED-TRAWL FISHING PROTOCOLS

The WT conducted normal annual surveys of both Div. The AN fished alongside the WT for comparison purposes only. Otherwise, the fishing protocols used were the same as in previous comparisons with these vessels (Warren et. al. 1997). On level bottom, the two vessels towed side by side at a intended distance of 0.5 nautical miles (nm) apart, or 0.9 km. Both vessels followed normal survey-fishing protocols, with tow durations of 15 minutes and tow speeds of 3 nm per hour. The vessels were instructed to tow on the same course, and the WT relayed the course to the AN. On slope edges, where side by side tows were not feasible due to depth differences, one vessel towed ahead of the other, alternating the lead vessel on a tow-by-tow basis. This was done so that the end of the tow for the trailing vessel occurred at a position just before the start of the tow for the leading vessel; that is, there was no overlap in the area covered by the 2 tows. The same depth range for each paired tow was maintained as close as possible between vessels. Differences in the depths fished for both vessels were minimized, for a target of less than 10% during comparative tows. If the WT had an unsuccessful set (e.g. torn gear), both vessels repeated their tows, moving slightly so that the same grounds were not towed over again.

Numbers and weights for all species caught were collected using the normal survey procedures. Biological sampling on the WT also followed survey protocols. Some biological sampling and collection of specimens could be switched to the AN, where possible, in order to avoid delays. Otherwise, no detailed biological sampling (otoliths, weight analysis, etc.) was required on the AN. Length measurements were collected on both vessels using the same survey protocols (e.g. length, sex, and maturity data where necessary), although the AN staff were able to measure more lengths because they were not required to do other biological sampling. Trawl-mounted CTD systems were used on both vessels, to measure water temperatures and to allow more precise post-survey analysis of time on bottom.

STATISTICAL MODELS

Let R_{ils} be the number of length l fish caught at the ith tow station by vessel s. We refer to the replacement vessel as $s = t$ for the test vessel, and we refer to the vessel to be replaced as $s = c$ for control. We assume that the replacement vessel is the AN, although our results can easily be adjusted if the WT is the replacement vessel. Let λ_{ils} denote the total standardized fish density for length l fish that vessel s encountered at tow station i. The total standardization is for area swept based on a tow distance of 0.8 nm and standard gear geometry.

Fixed effects model

We assume the probability a fish is captured, denoted as q_{ls}, is the same at each site i but possibly different for each vessel, s, and length, l, although we expect that q_{ls} varies smoothly in terms of l. The relative efficiency of the WT compared to the AN is defined as

$$\rho_l = \frac{q_{lc}}{q_{lt}}. \tag{2}$$

We also assume that each vessel encounters the same local fish densities; that is, $\lambda_{ilt} = \lambda_{ilc} = \lambda_{il}$ for all lengths l . If fish are captured independently of each other then the catch by the test vessel is a Poisson random variable with mean

$$E(R_{ilt}) = q_{lt}\lambda_{il} = \mu_{ilt}. \tag{3}$$

The catch by the control vessel is also a Poisson random variable with

$$E(R_{ilc}) = q_{lc}\lambda_{il} = \rho_l \mu_{ilt}, \tag{4}$$

For the Poisson distribution, $Var(R) = E(R)$.

The ρ_l's can be estimated using a Poisson generalized linear model (GLIM; e.g. McCullagh and Nelder 1989). This is essentially the approach used by Benoît and Swain (2003), although they adjusted for extra-Poisson variability, $Var(R) = \phi E(R)$. Let $R_{il} = R_{ilc} + R_{ilt}$ be the total catch-at-length for tow station i. If $R_{il} = 0$ then the two zero catches-at-length l from each vessel supply no information about ρ_l. Let n be the total number of $R_{il} > 0$ for all tow stations and lengths. There are $2n$ observations to estimate the n density parameters (i.e. μ's) and the ρ_l parameters. The Poisson GLIM approach is complicated because n can be very large if many tow stations and length classes are sampled (i.e. $n > 1000$), which means that there are many μ parameters to estimate. This also complicates constructing confidence intervals for ρ_l.

A better approach for inferences about ρ_l's (Section 4.5 in Cox and Snell 1989) is to use the conditional distribution of R_{ilc} given R_{il}. Reid (1995) provides considerable information and discussion on the role of conditioning in statistical inference. Let r_{il} be the observed value of R_{il}. The conditional distribution of R_{ilc} given $R_{il} = r_{il}$ is Binomial, with

$$\Pr(R_{ilc} = x | R_{il} = r_{il}) = \binom{r_{il}}{x} p_l^x (1 - p_l)^{r_{il} - x},$$

where $p_l = \rho_l/(1 + \rho_l)$ is the probability a captured fish is taken by the control (i.e. WT) vessel. The only unknown parameters in this distribution are the ρ_l's. The many μ nuisance parameters in (3) and (4) are eliminated in the conditional likelihood. There are n conditional observations to estimate the relative catchabilities. Note that standard first-order asymptotic inferences based on the Poisson or Binomial approaches are the same (Cox and Snell 1989); however, frequency distributions of estimators for the Binomial approach are based on idealized resampling in which the R_{il}'s are fixed, whereas this is not the case in the Poisson approach. More accurate inferences procedures, such as the bootstrap, may give different results for the two approaches, although we do not pursue this point further.

For the Binomial distribution $E(R_{ilc}) = r_{il}p_l$ and $Var(R_{ilc}) = r_{il}p_l(1 - p_l)$. An approach to deal with over-dispersion is to use a quasi-likelihood (McCullagh and Nelder 1989) with $Var(R_{ilc}) = \phi r_{il}p_l(1-p_l)$. Note that the over-dispersed Poisson approach may give difference statistical inferences (e.g. confidence intervals) than the over-dispersed Binomial approach. They are not equivalent. The Binomial approach seems preferable, for reasons outlined in Cox and Snell (1989) and Reid (1995).

We feel that ρ_l varies smoothly in terms of length, similar to q_{ls}. Relative efficiency is non-negative and in most situations will be a monotone function of length. In this case, a suitable and common parametric model is $\rho_l = \exp(\beta_o + \beta_1 l)$. This leads to the logistic regression model

$$p_l = \frac{\exp(\beta_o + \beta_1 l)}{1 + \exp(\beta_o + \beta_1 l)},$$

which is the canonical link function for the Binomial distribution (McCullagh and Nelder 1989). This model is commonly used in fishing gear size-selectivity studies (Millar 1992). In some situations the exponential model may not capture the smooth nature of ρ_l sufficiently (e.g. Millar et. al. 2004), and more complicated approaches may be required. We examine residual plots to check the lack-of-fit for the exponential model.

We refer to the logistic regression model approach as the FE2 model, for fixed effects model with two parameters. We also examine a model in which β_1 is fixed at zero and only an intercept or vessel effect parameter is estimated. We refer to this as the FE1 model. In this case we can also pool data over lengths because ρ_l is constant for all l. We refer to this as the FEP1 model. Pooling is a common approach to avoid complications that arise when within-set catches are not independent. We consider this point further in the section on mixed effects models. We use SAS/STAT® PROC GENMOD software to estimate the fixed effects models.

Subsampling and swept area adjustments

Further adjustments are required because of subsampling of the catch and because of variations in tow distance. The adjustments we use are similar to those considered by Millar (1994) and Cadigan et. al. (1996) for gear selectivity studies when good information on subsampling fractions was available. We present the adjustments in terms of the FE2 model, but the results are used in the same way with the mixed effects model presented in the next section. Let d_{is} be the distance towed in nautical miles (nm) for vessel s at tow station i, and let f_{ils} be the estimated sampling fraction, which we describe below. The right-hand sides of equations (3) and (4) need to be multiplied by $f_{ils}d_{is}/0.8$, where the targeted tow distance is 0.8 nm. It is easy to show in this case that

$$\log\left(\frac{p_{il}}{1 - p_{il}}\right) = \beta_o + \beta_1 l + \log\left(\frac{d_{ic}f_{ilc}}{d_{it}f_{ilt}}\right). \quad (5)$$

The last term can be treated as an offset in the logistic regression model (McCullagh and Nelder 1989). Note that the probability a fish is captured by the control vessel (p) depends on the tow station (i) if $d_{ic} \neq d_{it}$ or $f_{ilc} \neq f_{ilt}$.

An alternative approach is to scale the catches to estimate what they would be if all catches were sampled in a 0.8 nm tow. However, this produces artificial sample sizes and does not accurately reflect the information in zero catches. This is particularly important if no over-dispersion parameter is used, because standard errors will be too small and confidence intervals too narrow if the adjusted catch is increased substantially. The offset approach seems preferable, although it is not critical when an overdispersion parameter is used.

Catches are placed in baskets, and the number of baskets are subsampled. The subsampling fraction is estimated using the number of baskets sampled or the fraction of the total catch weight sampled, although clearly the actual fraction subsampled from each length class will vary from such estimates. In a small number of tows the catch is split (into small and large sizes) and subsampled differently, and the subsampling fractions for each split sample (referred to as 1 and 2) are recorded. However, the split may not be perfect and some small fish may be included with the large sized fish. In this situation we estimate the total fraction subsampled using the catches,

$$f_{ils} = \frac{r_{1ils} + r_{2ils}}{r_{1ils}/f_{1ils} + r_{2ils}/f_{2ils}}.$$

This does not work for length classes with $r_{1ils} = r_{2ils} = 0$. In this case we set $f_{ils} = (f_{1ils} + f_{2ils})/2$.

Differential subsampling also causes complications for modelling total or pooled catches within sets because the subsampling fractions depend on length, l. The expected pooled catches are

$$E(R_{it}) = \frac{d_{it}}{0.8} \sum_l f_{ilt}\mu_{ilt} \text{ and } E(R_{ic}) = \frac{\rho d_{ic}}{0.8} \sum_l f_{ilc}\mu_{ilt}.$$

If $f_{ilt} \neq f_{ilc}$ for some l then the conditional distribution of R_{ic} given R_i still depends on μ_{ilt}'s and μ_{ilc}'s, and this complicates the estimation of ρ.

The solution we use is to estimate a common subsampling fraction for each set and ignore within-set variations in subsampling. This is reasonable for our data because differential subsampling of catches occurred in only small numbers of sets (see Results). The subsampling fraction was estimated using the ratio of total catches to total raised catches,

$$f_{is} = \frac{\sum_l r_{ils}}{\sum_l r_{ils}/f_{ils}}.$$

However, using this pooled subsampling fraction means that a slightly different ρ estimate may be obtained using pooled and un-pooled catches.

Mixed effects model

In this approach we do not assume that $\lambda_{ilt} = \lambda_{ilc}$. Let $\delta_{il} = \log(\lambda_{ilc}/\lambda_{ilt})$ and let z_{il} denote the offset term in (5). The model for the WT proportion of catch is

$$\log\left(\frac{p_{il}}{1 - p_{il}}\right) = \beta_o + \beta_1 l + z_{il} + \delta_{il}. \quad (6)$$

If exactly the same length distributions of fish were encountered by both vessels then $\delta_{il} = 0$. In practise this does not happen. The length distributions can be substantially different, and the differences can vary systematically with length. For example, if the AN encountered larger fish in a tow compared to the WT then δ_{il} would decrease with length. This is illustrated in Fig. 1. In this hypothetical example δ_{il} decreases almost linearly with length. It is easy to construct examples where the change is not linear, but we suggest that in general δ_{il} will vary smoothly with length. However, we expect that for each length l the variations in δ_{il} will be independent between sites; that is, differences in length distributions should be uncorrelated across tow sites.

We use a mixed effects model to account for this error structure. A mixed model contains both fixed parameters and random "parameters", although parameters are usually considered to be fixed but unknown, so the random "parameters" are usually referred to as random effects. We assume that the δ's are random variables from a Normal distribution with mean zero, but the δ's are autocorrelated in terms of length; that is, $E(\delta_{il}) = 0$, $Var(\delta_{il}) = \sigma^2$ and $Corr(\delta_{ij}, \delta_{ik}) = \gamma^{|j-k|}$. This is an AR(1) correlation structure, with γ autocorrelation. The δ_{il} are assumed to be uncorrelated between sites; that is, $Corr(\delta_{ij}, \delta_{kl}) = 0$ for sites $i \neq k$ and for all lengths j, l. This model can account for smooth deviations from linearity in the logit proportion of total catch by the WT, caused by partly systematic differences in local stock densities fished by each vessel. The fixed effects are β_o and β_1. We do not use an additional over-dispersion parameter like in FE models because this variation will be captured by $Var(\delta_{il})$.

We use the new SAS/STAT PROC GLIMMIX software for estimation. PROC GLIMMIX software fits generalized linear mixed models (GLMM's) based on linearizations. A Taylor's series expansion is used to approximate the GLMM as a linear mixed model. The advantage of the linearization is that only the variance parameters have to be estimated numerically because closed form expressions exist for estimates of the regression parameters. Linearization fitting methods are doubly iterative. The approximate linear mixed model is fit, which is itself an iterative process, then the new parameter estimates are used to update the linearization, which results in a new linear mixed model. The process stops when parameter estimates between successive linear mixed model fits change within a specified tolerance only. The default estimation method in PROC GLIMMIX software for models containing random effects is a technique known as restricted pseudo-likelihood (see below) estimation with an expansion around the current estimate of the best linear unbiased predictors of the random effects. Maximum likelihood estimates of variance parameters tend to be biased for small sample sizes. The restricted pseudo-likelihood method may provide less biased estimation of random effect variance parameters.

An advantage of linearization-based methods is that they use a relatively simple form of the linearized model that typically can be fit based on only the mean and variance in the linearized form. Models for which the marginal distribution is difficult, or impossible, to compute can be fit with linearization-based approaches. The approach is well-suited for models with correlated errors, large number of random effects, crossed random effects, and multiple types of subjects.

Disadvantages of this approach are the absence of a true objective function for the overall optimization process and potentially biased estimates of the covariance parameters, especially for binary data. In a GLMM it is not always possible to derive the log likelihood of the data. Likelihood-based tests and statistics are often difficult to derive. PROC GLIMMIX software produces Wald-type test statistics and confidence intervals. We used the between-within method option to determine the denominator degrees of freedom for the fixed effects. We tested all the provided options in the pooled analysis and found no, or occasionally very small, differences in p-values.

PROC GLIMMIX software provides marginal and conditional residuals, on the data or link scale. Conditional residuals are based on predictors of the random effects and estimates of the fixed effects regression parameters. The predictors of random effects are the estimated best linear unbiased predictors (BLUPs) in the approximated linear model. We examine conditional Pearson-type residuals for goodness of fit. We sum raw residuals for all sets or lengths, depending on the focus of the residual analysis, and divide the summed residuals by their estimated standard deviations. We refer to such residuals as standardized residuals.

We also use SAS/STAT PROC NLMIXED software to examine the robustness of estimates to the estimation method; however, we can do this only for the more simple pooled models. Models with more complicated random effects such as those with auto-correlation are difficult to implement with PROC NLMIXED software, but straight-forward with PROC GLIMMIX software. PROC NLMIXED software fits nonlinear mixed models, including logistic regression, by maximizing an approximation to the likelihood integrated over the random effects, which is different than PROC GLIMMIX software. Such marginal estimation methods are commonly used with mixed effects models. PROC NLMIXED software only implements maximum likelihood. This is because the analog to the restricted maximum likelihood method in PROC NLMIXED software would involve a high dimensional integral over all of the fixed-effects parameters, and this integral is typically not available in closed form.

Similar to the fixed effects models, we denote the mixed effects model with only a vessel effect in (6), e.g. $\beta_1 = 0$, as ME1, and we refer to the mixed effects model with an intercept and length parameter as ME2. If we pool data then we denote the method as MEP1.

RESULTS

During the 2005 spring survey in 3Ps, a total of 66 successful paired tows were completed involving the WT and AN. Additional sets were planned in 3LNO, but due to timing and mechanical problems, this work was not attempted. During the fall of 2005, additional comparative fishing (CF) between the WT and AN was carried out, directed at species and depths that had little coverage in the spring 3Ps CF. This resulted in an additional 40 paired sets, 14 in 3N and 26 in 3L. The 3L CF was directed mainly at crab and shrimp, and the 3N CF at shallow-water species, primarily yellowtail flounder. Overall, 106 paired survey sets were carried out in 2005. Most major commercial species had some survey coverage, although there were few sets in the 500-730 m range (the deepest areas WT would cover in spring and fall surveys), so deepwater species such as Greenland halibut may be data deficient.

Sets were located in the far offshore portion of 3Ps (Fig. 2), the shelf area in 3N, and the northern part of 3L (Fig. 3). The distance between paired tows was relatively constant, with a maximum of 3.7 km for set 75 in 3Ps (Fig. 2). The set numbering is unique within NAFO Div. Tow depths were also usually similar (Fig. 4), with a maximum absolute difference of 37 m for set 75 in 3Ps. This set has potential to result in catch outliers.

Seven species of fish (Table 1) were selected to assess the relative efficiency of the WT. Species of crab and shrimp were also measured but are not considered here. More fish were measured on the AN than on the WT. Standardized differences in scaled catches, $r^* = 0.8r/(d \times f)$, were examined for potential outliers (Fig. 5-11). The standardization was based on Poisson variability,

$$\frac{r^*_{ic} - r^*_{it}}{(r^*_{ic} + r^*_{it})^{1/2}}.$$

We use the notation $r_{is} = \sum_l r_{ils}$. These differences do not account for over-dispersion; hence, too many pairs will be identified as outliers. We do not use this approach to identify outliers, but simply to assist in understanding the within-pair variability in catches.

Two large differences in catches occurred for cod (Fig. 6) in the western portion of 3Ps. The catch weights (not shown) for these sets were also substantially different. Model residuals for these sets are examined later. A larger number of sets for redfish are potentially outliers (Fig. 7). yellowtail flounder were observed in relatively few sets, mostly in 3N, and three of these paired sets had substantial differences catches (Fig. 11). The catch weights were more similar for these sets, suggesting that the differences may involve small fish.

FIXED EFFECTS MODEL (FE1)

In this model the length parameter in (5) is fixed at zero and the intercept (β_o) is treated as an unknown fixed quantity to estimate. In the first analysis catches were pooled over length classes within each set (FEP1). The number of set-pairs where differential subsampling occurred is usually low, except for Deepwater redfish (Table 2). Five of seven estimates were negative (Table 2) indicating that the WT had a slightly lower relative efficiency than the AN, although only witch flounder had a relative efficiency that was significantly different from one. If catches were not pooled then four of seven stocks had β_o's that were significantly different from zero (Table 3). Note that $\hat{\beta}_o$ for Atlantic cod was identical from pooled and un-pooled catches because there was no differential subsampling for this species. The FE1 model results may be too liberal because of the potential lack of independence of catches within sets. The largest difference in $\hat{\beta}_o$'s was for Deepwater redfish which was the species with the most differential subsampling.

Total scaled catches (for all lengths) from each vessel are shown in Fig. 12 and 13. The estimated relative efficiency from the FEP1 model is also shown (dashed line) as a line through the origin with slope $\hat{\rho} = \exp(\hat{\beta}_o)$. The total scaled catch (for all sets) by the AN was somewhat greater than by the WT for five of seven stocks. However, the $\hat{\rho}$'s for the seven species do not appear substantially different from one. Some potential outliers are apparent for Atlantic cod, deepwater redfish, thorny skate, and witch flounder. Sets with absolute deviance residuals larger than three were identified with their set number.

Another approach mentioned in the Methods was to estimate the FE1 model based on catches raised by the subsampling fraction, and also standardized to a 0.8 nm tow distance. In this approach there is no offset. Five stocks had significant vessel effects (Table 4) based on this approach. Estimates were substantially different compared to the above offset approaches. For example, the estimate of β_o for Atlantic cod from the FE1 and FEP1 models was 0.1553, but the analogous estimate based on raised catches was 0.5167. The $\hat{\rho}$ values were 1.167 & 1.67, respectively. The effect for yellowtail flounder was significant based on raised catches, whereas it was not significant based on offset analyses (Tables 2 and 3).

MIXED EFFECTS MODEL (ME1)

The random effects were also pooled or summed in MEP1, and $\delta_i = \sum_l \delta_{il}$ were assumed to be independent and identically distributed (iid) $N(0, \sigma^2),\ i = 1, 2, \ldots$. Five of seven β_o estimates were negative (Table 5), again indicating that the WT had a slightly lower relative efficiency than the AN. The effects were not significant for any species. Note that $\hat{\beta}_o$ for some species had a different sign compared to the FEP1 results (Table 2), and there was not a close correspondence between the fixed and random effects estimates. MEP1 model estimates of ρ (solid line; Fig. 12 and 13) were close to one. The predicted random effects were usually close to zero with some exceptions.

In the first ME1 model we assumed the δ_{il}'s were constant across lengths for each site and equal to δ_i which were independent and identically distributed (iid) $N(0, \sigma^2)$ for $i = 1, 2, \ldots$. This makes the ME1 model more comparable with the MEP1 model. The un-pooled results (Table 6), including estimates, standard errors, and confidence intervals, were similar to the pooled results. The Atlantic cod results were identical. In the fixed effects situation the estimates were similar but the standard errors and confidence intervals were very different.

Estimates and confidence intervals based on PROC NLMIXED software (Table 7) were similar to those obtained using PROC GLIMMIX software (Table 5). The largest discrepancies occurred with Atlantic cod.

If, instead, the random effects were modelled as autocorrelated random variables as outlined in the Methods then the results (Table 8) differed somewhat, but none of the effects was significantly different from zero. In this approach the δ_{il}'s were not assumed to be perfectly correlated ($\gamma = 1$) as in the first ME1 model. The autocorrelation estimates ($\hat{\gamma}$) were greater than 0.9 which indicates that there were substantial length dependencies in the within-set catches. We consider this approach in more detail later.

OUTLIERS I

The FEP1 and MEP1 models were re-estimated after some of the potential outliers were removed. We removed two sets (i.e. trawl pairs) for each species. We chose to remove two potential outliers per stock to illustrate the outlier-sensitivity of the methods, and not because each data set had exactly two outliers. The sets are indicated in Fig. 12 and 13, except for Greenland halibut (set 16 in fall survey; set 100 in spring survey) and yellowtail flounder (sets 58 and 59 in fall survey). Residuals for these latter sets were relatively large, although not large enough to identify with set numbers (i.e. $\gtrless \pm 3$). MEP1 estimates of β_o appeared somewhat more stable (Fig. 14). They tended to not change as much when outliers were removed compared to FEP1 results. This was especially the case for Atlantic cod, where one set (116; Fig. 12) was more anomalous than the others. For the other stocks the two outliers, when large, tended to be opposite in sign which might tend to cancel their effect after removal. The MEP1 estimates were more sensitive than the FEP1 estimates for three stocks (American plaice, Greenland halibut, and yellowtail flounder), and the differences were very small. We further consider the outlier-robustness below.

The vessel effect was not significant in any of the MEP1 models with two potential outliers removed (table not presented), similar to the results based on all the data. The effects were significant for Atlantic cod and witch flounder in the FEP1 model with outliers removed (table not presented), whereas in the model based on all data (Table 2) only the witch flounder vessel effect was significant.

These results suggest that the random site effects approach may better accommodate outliers. While a rigorous examination of this is beyond the scope of this paper, it is informative to examine the fixed and mixed effects approaches when a single large outlier is added. This was done for Greenland halibut, whose observed data did not contain large outliers (Fig. 12). An outlier was added by multiplying by 20 the WT catch for set 105. Analyses were based on pooled data. This outlier (Fig. 15) was not unrealistic compared to some of the other species (e.g. Atlantic cod in Fig. 12). The MEP1 estimate $\hat{\beta}_o = 0.04781$ was affected much less by the outlier than the FEP1 estimate $\hat{\beta}_o = 0.8758$. In FEP1, β_o was significantly different from zero (p-value $<$ 0.0001) whereas in MEP1 β_o was not significant (p-value $=$ 0.7164).

Although the total WT catch with the outlier is over double the AN catch, the FEP1 estimate of ρ was not consistent with the majority of sets (Fig. 15; dashed line), and only 11 of 56 sets had relative efficiencies greater than the estimated value. The MEP1 estimate of ρ was more consistent, with 25 of 56 sets having relative efficiencies greater than the estimated value. The sign and signed rank tests that the mean of the chi-square residuals was zero from FEP1 were both significant (p-values < 0.0001), whereas they were not significant for MEP1 (p-values $>$ 0.5).

FIXED EFFECTS MODEL (FE2)

In this model (FE2) both parameters in (5) are treated as unknown fixed quantities to estimate, based of course on the un-pooled data. Parameter estimates and confidence intervals are presented in Table 9. Significant differences in relative efficiency were found in three of seven stocks. Estimated ρ_l's and model estimates of the proportion of the total catch-at-length from both vessels taken by the WT are shown in Fig. 16 and 17. The estimated ρ_l's were usually less than one and decreased with length for six of seven stocks, suggesting that the WT had a lower catchability than the AN. The observed proportions of catch by the WT were quite variable around the model predictions, although the sample sizes are not reflected in the proportions and would tend to be low for small and large lengths.

The length distributions of total catches and model residuals are shown in Fig. 18-24. The differences in length frequencies for the WT and AN tended to be small. The largest differences occurred for cod (Fig. 19), and were associated with differences in a small number of tows. Potential outliers were apparent, similar to the FE1 analysis. Some of the residuals-at-length also deviated substantially from zero (Fig. 18, 22, and 23) suggesting potential lack-of-fit of the logistic model for the proportion of total catch by the WT, although this may also be caused by outliers.

RANDOM EFFECTS MODEL (ME2)

We investigate two assumptions about the random effects. The first formulation is appropriate when the length distributions of fish encountered by each vessel are not equal but their log ratio varies smoothly in terms of length. The second formulation is more restrictive and is suitable when the ratios are constant within sites, but vary randomly between sites.

In the first ME2 model the random effects in (6) were modelled as autocorrelated random variables. A marginally significant length effect was found for deepwater redfish (Table 10), and there was some evidence of a difference in relative efficiency for American plaice. The estimates of ρ_l were considerably different from the FE2 model (Fig. 16 and 17) for some species.

The autocorrelation estimates ($\hat{\gamma}$; Table 10) were greater than 0.9 which indicates that there were substantial length dependencies in the within-set proportion of total catch by the WT, beyond that accounted for by the fixed length effect β_1. The predicted random effects deviated substantially from zero which suggests the ratio of fish densities for some lengths and sets were substantially different from one (Fig. 25-31). No large outliers were apparent from the set-total conditional ME2 chi-square residuals. Outliers appear to have been accounted for by the random effects. Standardized residuals were improved compared to those from the fixed effects model, although some length trends remained (Fig. 30).

The second ME2 model considered was based on a random site effect that was constant within sets. This is commonly referred to as a random intercept (RI) model. The results (Table 11) were considerably different from the FE2 model (Table 9) and first ME2 AR model (Table 10) results. The same species had significant effects as in the FE2 model, but the estimates of the effects were different for Atlantic cod and deepwater redfish. The fit of the ME2 RI model was substantially worse than the fit of the ME2 AR model for all stocks (Table 12). The fits were most similar for Greenland halibut. The ME2 AR model seems to be preferable because of the better fit.

OUTLIERS II

The FE2 and ME2 AR models were re-estimated after two potential outliers were removed. The sets removed were described in Outliers I. Estimates of β_o and β_1 for Atlantic cod appeared to be more stable from the random site effects model (Fig. 32). For the other stocks, both approaches seemed equally sensitive to the removal of the potential outliers.

SCANMAR SWEPT-AREA ADJUSTMENTS

We repeated the analyses using tow-specific swept-area calculations based on Scanmar trawl wing-spread measurements and estimated distances for each tow. The swept-areas from WT and AN tows replace the distance towed terms in (5). The offset terms only depend on the ratio of swept-areas for the WT and AN sets, and not their actual values.

Summary Scanmar statistics are presented in Table 13. The summaries are over all hauls within seasons. Only tow-pairs where both sensors were working are included in the summaries. These data will be considered in more detail in Walsh (2006).

Parameter estimates and confidence intervals are presented in Tables 14 and 15 for the FE2 and ME2 AR models. They are fairly similar (Fig. 33) to the estimates in Tables 9 and 10. Conclusions based on p-values are almost always the same with and without swept-area adjustments.

DISCUSSION

The results overall suggest that there were no significant differences in catchability between the WT and AN fishing the Campelen 1800 survey trawl. The WT and AN total length frequencies were similar, especially in light of the between-tow variability in catches. The sign of vessel effects for different species tended to be negative which may provide some additional evidence that the catchability of the WT was lower than the AN. However, the effect, if it exists, appeared small and could be ignored without serious consequences.

The more traditional fixed effects model suggested that some of the differences between vessels were significant. However, this model was based on an erroneous assumption that differences in stock densities fished by each trawler were either identical, which leads to Binomial variability, or completely random and independent of length, which leads approximately to over-dispersed Binomial variability. We conjecture that this was the motivation by Benoît and Swain (2003) and Lewy et. al. (2004) for using a Binomial over-dispersion parameter. We also analyzed the data using a mixed effects model with an independent and identically distributed random normal intercept for each set and length, and essentially found the same results as the fixed effects model. That is, we found statistically significant differences for the same three stocks.

We suggest that differences in stock densities will not be completely random. Rather, they will vary smoothly as a function of length. When this random structure is accounted for by using a length autocorrelated random component in the logistic model for the proportion of total catch by the WT, relative efficiency is significantly different from one for only one stock, and even then the differences in catch-at-length were not compelling. We suggest these mixed effects results are more reliable because the basis for statistical inference is more reliable. We also examined a random intercept model approach but found it did not fit our data nearly as well as the autocorrelated model. However, the efficacy of the mixed effects model approach for estimating relative efficiency and determining statistical significance requires further evaluation. Simulations would be useful for this purpose.

Another source of variability that we have not explicitly accounted for is between-set variability in catchability. This type of variability is commonly observed in cover-codend experiments (e.g. Fryer 1991, Millar et. al. 2004) that directly measure catchability. This will also produce between-site variability in ρ. We demonstrate this in Fig. 34. We generated 10 random examples of catchability curves (q_l), two for each pair (c and t) for five tow sites. The curves were generated from a logistic model with a 50% retention length (L_{50}) at 40 cm and a selection range $(L_{75} - L_{25})$ of 20 cm. The L_{50}'s were randomly generated from a normal distribution with a 10% coefficient of variation. The log of $\rho_l = q_{lc}/q_{lt}$ varied smoothly versus length for each pair, similar to what we expect with local spatial variability in stock densities (e.g. Fig. 1). These two sources of variability are confounded in paired-trawl experiments and the random effects we modelled represent the cumulative effects of both types of variability, although we have interpreted them as local spatial variability.

Heagerty and Kurland (2001) examined the impacts of using a random intercepts approach with correlated random effects. They demonstrated that large biases can occur when the wrong random effects assumptions are used.

We also presented evidence that a mixed effects model can better accommodate paired-trawl outliers. This is an advantage because identifying outliers and making appropriate adjustments for them can be an iterative and time consuming process. The mixed effects approach adjusts for outliers internally during estimation, and also adjusts inferences to account for outliers. This approach treats outliers of reasonable size as extreme values of the random effects. The fixed effects approach treats all data equally in estimation and only adjusts for the outlier-induced over-dispersion after parameter estimation. In fact, for the FE1 and FEP1 models, if there were no offsets to adjust for then the estimate of β_o would simply be $\log\{\hat{p}/(1-\hat{p})\}$, where $\hat{p}$ is the proportion of total catch by the WT for all sets. Clearly all sets get equal weight in this estimator. When subsampling is taken into account using an offset, which we feel is preferable, then the estimator of β_o in the FE1 and FEP1 models is not as simple and it does not have a closed form solution. However, our research into the impact of outliers was very limited, and more is required.

Subsampling of catches is necessary on research surveys when sampling time between sets is limited and/or the catch is large. A common approach when analyzing subsampled catches is to raise the catches by the subsampling fraction and treat these catches as observations. If an overdispersion parameter is used then the standard errors based on the raised catches may still be reasonable. However, estimates based on raised catches may be considerably different than estimates from the offset approach. The Atlantic cod data provide a good example of this. The estimate of β_o from the FE1 and FEP1 models was 0.1553, but the corresponding estimate based on raised catches was 0.5167. This estimate can be obtained as the logit of the proportion of total catch by the WT using the raised catches in Fig. 12. The offset approach suggests that the WT is 17% more efficient than the AN (i.e. $\hat{\rho} = 1.168$) whereas the raised-catch approach suggests that the WT is 68% more efficient (i.e. $\hat{\rho} = 1.676$). These are very different estimates. Raised catch results for thorny skate and yellowtail flounder were also substantially different, and they suggested a significant vessel effect for yellowtail flounder. The offset approaches did not. We feel the offset approach is the more appropriate way to account for subsampling; however, further research to validate this would be useful. This should include the situation when parts of the catch are subsampled differently.

Pooling of catches within sets may be advisable for fixed effects models when catches within some factor levels (e.g. set or site) are correlated. This correlation can be directly accommodated in a mixed effects model, and pooling does not seem necessary with this approach. We demonstrated for Atlantic cod (Tables 5 and 6) that identical estimates and inferences can be obtained from pooled and un-pooled mixed effects models. This is advantageous because, as described in the Methods section, pooling can be problematic when catches are subsampled differently so that subsampling fractions are length-dependent. However, the are many options in PROC GLIMMIX to specify degrees of freedom for statistical inferences about fixed effects, and the best approach for calibration studies requires more investigation.

We did not find that adjustments for tow-specific swept-area estimates changed estimates of relative efficiency, or improved precision. This would be expected if swept-area does not vary much between tows or if the estimate of swept area was too imprecise. It would also be expected if catches were not affected by swept-area within the range of variation in swept-areas that occurred in the comparative fishing experiments.

Cotter (2001) estimated survey vessel calibration factors based only on a simple year class model, and did not utilize any comparative fishing information. While this is possible to do, the year class model will in most circumstances not provide reliable estimates of cohort numbers-at–age. The logical way to improve Cotter's (2001) approach is to estimate calibration factors using a more realistic cohort model such as VPA; however, this clarifies the fundamental problem with the method. When a survey vessel is replaced then Cotter's (2001) approach is essentially the same as estimating different catchabilities. This breaks the convergence property of the VPA, and the survey time series subsequently becomes much less informative about current stock size. It is inappropriate to estimate calibration factors with a year class model and then apply the calibration factors in a VPA without an analytical acknowledgement that the same data have been used twice. Such an analytical acknowledgement could take the form that older converted survey indices are correlated with recent indices obtained with the new survey vessel.

Wilderbuer and Kappenman (1998) examined some linear model approaches with log-transformed catches and another less parametric approach. They noted that treating stock densities as completely different between different pairs of tow stations may be inefficient. A better approach may be to stratify and treat stock densities at different tow sites within-strata as somehow correlated. Such approaches may improve statistical power, but require further investigation. However, the within-pair variability we observed in 3Ps and 3LN suggest that potential improvements due to stratification or a some other spatial modelling may be limited.

REFERENCES

Benoît, H.P., and Swain, D.P. 2003. Accounting for length-and depth-dependent diel variation in catchability of fish and invertebrates in an annual bottom-trawl survey. ICES J. Mar. Sci. 60: 1298-1317.

Cadigan, N.G., Boulos, D.L. and Hickey, W.M. 1996. Analysis of subsampled catches from trouser trawl size selectivity studies. J. Northw. Atl. Fish. Sci. 19: 41-49.

Cotter, A.J.R. 2001. Intercalibration of North Sea International Bottom Trawl Surveys by fitting year-class curves. ICES J. Mar. Sci. 58: 622-632.

Cox, D.R., and Snell, E.J. 1989. Analysis of Binary Data. 2nd ed. Chapman and Hall, London.

Doubleday, W.G. [Ed.] 1981. Manual on groundfish surveys in the Northwest Atlantic. Northwest Atlantic Fisheries Organization Scientific Council Studies, 2: 7-55.

Fanning, L.P. 1985. Intercalibration of research survey results obtained by different vessels. CAFSAC (Canadian Atlantic Fisheries Scientific Advisory Committee) Research Document 85/3.

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

Heagerty, P.J., and Kurland, B.F. 2001. Misspecified maximum likelihood estimates and generalised linear mixed models. Biometrika 88: 973-985.

Lewy, P., Nielsen, J.R., and Hovgård, H. 2004. Survey gear calibration independent of spatial fish distribution. Can. J. Fish. Aquat. Sci. 61: 636-647.

McCallum, B., and Walsh, S.J. 2002. An update on the performance of the Campelen 1800 during bottom trawl surveys in NAFO Subareas 2 and 3 in 2001. NAFO SCR Doc. 02/32:16p

McCullagh, P., and Nelder, J.A. 1989. Generalized linear models. 2nd ed. Chapman and Hall, London.

McCulloch, C.E., and Searle, S.R. 2001. Generalized, linear, and mixed models. John Wiley & Sons, New York.

Millar, R.B. 1992. Estimating the size-selectivity of fishing gear by conditioning on the total catch. J. Amer. Stat. Assoc. 87**:** 962-968.

Millar, R.B. 1994. Sampling from trawl gears used in size-selectivity experiments. ICES J. Mar. Sci. 51: 293-298.

Millar, R.B, Broadhurst M.K., and Macbeth, W.G. 2004. Modelling between-haul variability in the size-selectivity of trawls. Fish. Res. 67: 171-181.

Pelletier, D. 1998. Intercalibration of research survey vessels in fisheries: a review and an application. Can. J. Fish. Aquat. Sci. 55: 2672-2690.

Reid, N. 1995. The roles of conditioning in inference. Stat. Sci. **10**: 138-199.

Smedbol, R.K., Shelton, P.A., Swain, D.P., Frechet, A., and Chouinard, G.A. 2002. Review of population structure, distribution and abundance of cod (*Gadus morhua*) in Atlantic Canada in a species-at-risk context. DFO Can. Sci. Advis. Sec. Res. Doc. 2002/082.

Walsh, S.J. 2006. Effect of vessels and crew on geometry and performance of Campelen 1800 shrimp trawl during comparative fishing in the Newfoundland region in the spring and fall of 2005. DFO Can. Sci. Advis. Sec. Res. Doc. To appear.

Warren, W., Brodie, W., Stansbury, D., Walsh, S., Morgan, J., and Orr, D. 1997. Analysis of the 1996 comparative fishing trial between the *Alfred Needler* with the Engel 145 trawl and the *Wilfred Templeman* with the Campelen 1800 trawl. NAFO SCR Doc. 97/68.

Wilderbuer, T.K., and Kappenman, R.F. 1998. Analysis of Fishing Power Correction Factor Estimates from a Trawl Comparison Experiment. N. Am. J. Fish. Manag. 18: 11-18.

Table 1. Catch summaries. R_c and R_t are the WT and AN measured catches. n is the total number of observations (lengths and tows) where $R_t + R_c > 0$.

Species	Scientific Name	n	R_c	$R_t + R_c$	R_t/R_c
American plaice	*Hippoglossoides platessoides*	2035	5603	11494	1.051
Atlantic Cod	*Gadus morhua*	1132	1899	3926	1.067
deep water redfish	*Sebastes mentella*	1030	5469	12069	1.207
Greenland halibut	*Reinhardtius hippoglossoides*	585	606	1359	1.243
thorny skate	*Raja radiata*	990	1127	2394	1.124
witch flounder	*Glyptocephalus cynoglossus*	970	2162	5046	1.334
yellowtail flounder	*Limanda ferruginea*	536	2576	5795	1.250

Table 2. FEP1 model results. Se - standard error. L,U - profile likelihood confidence intervals. pv is the χ^2 p-value. Diff f is the number of set-pairs with differential subsampling between lengths. Significant estimates in bold.

Species	$\hat{\phi}$	$\hat{\beta}_o$	Se	95% L	95% U	χ^2	pv	Diff. f
American plaice	2.99	0.0188	0.0561	-0.0912	0.1286	0.11	0.7379	3
Atlantic Cod	4.34	0.1553	0.1432	-0.1258	0.4362	1.18	0.2779	0
deep water redfish	5.93	-0.1156	0.1118	-0.3357	0.1031	1.07	0.3015	9
Greenland halibut	1.33	-0.0803	0.0727	-0.2233	0.0619	1.22	0.2695	1
thorny skate	2.42	-0.1383	0.0995	-0.3339	0.0566	1.93	0.1647	2
witch flounder	2.43	**-0.1533**	0.0693	-0.2896	-0.0178	4.89	0.0270	2
yellowtail flounder	3.48	-0.0249	0.0923	-0.2065	0.1554	0.07	0.7870	1

Table 3. FE1 model results. Se - standard error. L,U - profile likelihood confidence intervals. pv is the χ^2 p-value. Significant estimates in bold.

Species	$\hat{\phi}$	$\hat{\beta}_o$	Se	95% L	95% U	χ^2	pv
American plaice	1.37	0.0144	0.0257	-0.0360	0.0647	0.31	0.5758
Atlantic Cod	1.76	**0.1553**	0.0579	0.0419	0.2687	7.21	0.0073
deep water redfish	2.02	**-0.1587**	0.0382	-0.2337	-0.0840	17.28	$<$.0001
Greenland halibut	1.18	-0.0812	0.0644	-0.2077	0.0447	1.59	0.2069
thorny skate	1.36	**-0.1411**	0.0560	-0.2509	-0.0315	6.36	0.0117
witch flounder	1.28	**-0.1611**	0.0365	-0.2327	-0.0898	19.53	$<$.0001
yellowtail flounder	1.38	-0.0291	0.0366	-0.1010	0.0425	0.63	0.4261

Table 4. FE1 model results from raised catches. Se - standard error. L,U - profile likelihood confidence intervals. pv is the χ^2 p-value. Significant estimates in bold.

Species	$\hat{\phi}$	$\hat{\beta}_o$	Se	95% L	95% U	χ^2	pv
American plaice	1.39	0.0097	0.0250	-0.0393	0.0588	0.15	0.6974
Atlantic Cod	2.00	**0.5167**	0.0546	0.4100	0.6243	89.38	<.0001
deep water redfish	2.64	**-0.1752**	0.0322	-0.2384	-0.1121	29.57	<.0001
Greenland halibut	1.13	-0.0795	0.0645	-0.2061	0.0469	1.52	0.2179
thorny skate	1.38	**-0.2768**	0.0531	-0.3812	-0.1729	27.12	<.0001
witch flounder	1.28	**-0.1714**	0.0367	-0.2434	-0.0996	21.83	<.0001
yellowtail flounder	2.78	**-0.0911**	0.0339	-0.1575	-0.0246	7.22	0.0072

Table 5. MEP1 model results. Se - standard error. L,U - Wald confidence intervals. pv is the t-statistic p-value.

Species	$\hat{\sigma}^2$	$\hat{\beta}_o$	Se	95% L	95% U	t	pv
American plaice	0.47	0.05778	0.07921	-0.09930	0.2148	0.73	0.4674
Atlantic Cod	0.92	-0.05877	0.1344	-0.3259	0.2083	-0.44	0.6630
deep water redfish	0.85	-0.06005	0.1290	-0.3179	0.1978	-0.47	0.6432
Greenland halibut	0.10	-0.07765	0.07929	-0.2365	0.08125	-0.98	0.3317
thorny skate	0.50	0.02779	0.1135	-0.1982	0.2538	0.24	0.8073
witch flounder	0.30	-0.09396	0.09257	-0.2794	0.09147	-1.02	0.3144
yellowtail flounder	0.38	-0.01594	0.1512	-0.3288	0.2969	-0.11	0.9170

Table 6. ME1 model results. Se - standard error. L,U - Wald confidence intervals. pv is the t-statistic p-value.

Species	$\hat{\sigma}^2$	$\hat{\beta}_o$	Se	95% L	95% U	t	pv
American plaice	0.46	0.05520	0.07889	-0.1012	0.2116	0.70	0.4856
Atlantic Cod	0.92	-0.05877	0.1344	-0.3259	0.2083	-0.44	0.6630
deep water redfish	0.85	-0.08755	0.1292	-0.3458	0.1707	-0.68	0.5005
Greenland halibut	0.10	-0.07956	0.07910	-0.2381	0.07896	-1.01	0.3189
thorny skate	0.51	0.02428	0.1136	-0.2018	0.2504	0.21	0.8313
witch flounder	0.29	-0.1013	0.09150	-0.2846	0.08203	-1.11	0.2731
yellowtail flounder	0.38	-0.01881	0.1512	-0.3316	0.2939	-0.12	0.9020

Table 7. MEP1 model results using PROC NLMIXED software. Se - standard error. L,U - Wald confidence intervals. pv is the t-statistic p-value.

Species	$\hat{\sigma}^2$	$\hat{\beta}_o$	Se	95% L	95% U	t	pv
American plaice	0.49	0.05575	0.08097	-0.1048	0.2163	0.69	0.4926
Atlantic Cod	0.99	-0.07010	0.1405	-0.3491	0.2089	-0.50	0.6189
deep water redfish	0.86	-0.06420	0.1304	-0.3249	0.1965	-0.49	0.6243
Greenland halibut	0.10	-0.08090	0.07912	-0.2395	0.07766	-1.02	0.3110
thorny skate	0.54	0.02487	0.1173	-0.2086	0.2584	0.21	0.8326
witch flounder	0.30	-0.09702	0.09302	-0.2834	0.08932	-1.04	0.3014
yellowtail flounder	0.36	-0.01720	0.1490	-0.3255	0.2911	-0.12	0.9091

Table 8. ME1 AR model results. Se - standard error. L,U - Wald confidence intervals. pv is the t-statistic p-value.

Species	$\hat{\sigma}^2$	$\hat{\gamma}$	$\hat{\beta}_o$	Se	95% L	95% U	t	pv
American plaice	0.6314	0.9575	0.06350	0.07594	-0.08709	0.2141	0.84	0.4050
Atlantic Cod	1.6125	0.9739	-0.1006	0.1446	-0.3880	0.1867	-0.70	0.4884
deep water redfish	1.0743	0.9302	-0.1536	0.1164	-0.3863	0.07905	-1.32	0.1918
Greenland halibut	0.1353	0.9544	-0.07243	0.07958	-0.2319	0.08705	-0.91	0.3667
thorny skate	0.9652	0.9057	-0.02408	0.1002	-0.2236	0.1754	-0.24	0.8107
witch flounder	0.3875	0.9659	-0.05001	0.09207	-0.2345	0.1344	-0.54	0.5892
yellowtail flounder	0.4818	0.9630	-0.02355	0.1436	-0.3205	0.2734	-0.16	0.8712

Table 9. FE2 model results. Se - standard error. L,U - profile likelihood confidence intervals. pv is the χ^2 p-value. Significant estimates in bold.

Species	$\hat{\phi}$		Estimate	Se	95% L	95% U	χ^2	pv
American plaice	1.36	β_o	**0.2297**	0.0689	0.0947	0.3648	11.11	0.0009
		β_1	**-0.0088**	0.0026	-0.0139	-0.0037	11.31	0.0008
Atlantic Cod	1.75	β_o	**0.6112**	0.1485	0.3211	0.9036	16.94	$<$.0001
		β_1	**-0.0119**	0.0029	-0.0155	-0.0040	11.07	0.0009
deep water redfish	2.00	β_o	**0.2711**	0.1043	0.0667	0.4756	6.76	0.0093
		β_1	**-0.0198**	0.0045	-0.0286	-0.0110	19.42	$<$.0001
Greenland halibut	1.18	β_o	-0.1228	0.3305	-0.7733	0.5244	0.14	0.7103
		β_1	0.0011	0.0083	-0.0152	0.0174	0.02	0.8980
thorny skate	1.36	β_o	-0.0187	0.1198	-0.2534	0.2164	0.02	0.8761
		β_1	-0.0031	0.0027	-0.0083	0.0021	1.33	0.2481
witch flounder	1.28	β_o	-0.0390	0.1246	-0.2836	0.2048	0.10	0.7544
		β_1	-0.0044	0.0043	-0.0128	0.0040	1.05	0.3054
yellowtail flounder	1.37	β_o	0.1785	0.1235	-0.0637	0.4205	2.09	0.1483
		β_1	-0.0074	0.0042	-0.0156	0.0008	3.09	0.0786

Table 10. ME2 AR model results. Se - standard error. L,U - Wald confidence intervals. pv is the t-statistic p-value. Significant estimates in bold.

Species	$\hat{\sigma}^2/\hat{\gamma}$		Estimate	Se	95% L	95% U	t	pv
American plaice	0.6217	β_o	0.2835	0.1502	-0.01435	0.5813	1.89	0.0619
	0.9568	β_1	-0.00741	0.004382	-0.01601	0.001181	-1.69	0.0909
Atlantic Cod	1.6322	β_o	-0.1663	0.2701	-0.7030	0.3704	-0.62	0.5397
	0.9741	β_1	0.001419	0.004914	-0.00822	0.01106	0.29	0.7728
deep water redfish	1.0562	β_o	0.2275	0.2274	-0.2271	0.6821	1.00	0.3210
	0.9309	β_1	-0.01667	0.008586	-0.03352	0.000177	-1.94	0.0525
Greenland halibut	0.1464	β_o	0.1201	0.3505	-0.5822	0.8225	0.34	0.7331
	0.9539	β_1	-0.00486	0.008655	-0.02186	0.01214	-0.56	0.5747
thorny skate	0.9736	β_o	0.02653	0.2577	-0.4865	0.5395	0.10	0.9183
	0.9065	β_1	-0.00099	0.004712	-0.01024	0.008254	-0.21	0.8329
witch flounder	0.3922	β_o	-0.1141	0.1981	-0.5109	0.2827	-0.58	0.5668
	0.9659	β_1	0.002248	0.006093	-0.00971	0.01421	0.37	0.7122
yellowtail flounder	0.4894	β_o	0.1779	0.2713	-0.3834	0.7392	0.66	0.5185
	0.9633	β_1	-0.00693	0.007910	-0.02247	0.008606	-0.88	0.3811

Table 11. ME2 random intercept model results. Se - standard error. L,U - Wald confidence intervals. pv is the t-statistic p-value. Significant estimates in bold.

Species	$\hat{\sigma}^2$		Estimate	Se	95% L	95% U	t	pv
American plaice	0.4371	β_o	**0.2667**	0.1040	0.06050	0.4729	2.56	0.0118
		β_1	**-0.00755**	0.002496	-0.01245	-0.00266	-3.03	0.0025
Atlantic Cod	0.9509	β_o	**-0.3555**	0.1733	-0.6997	-0.01121	-2.05	0.0431
		β_1	**0.007028**	0.002535	0.002055	0.01200	2.77	0.0057
deep water redfish	0.8745	β_o	**0.6823**	0.1536	0.3752	0.9895	4.44	$<$.0001
		β_1	**-0.03545**	0.003729	-0.04277	-0.02813	-9.51	$<$.0001
Greenland halibut	0.1028	β_o	0.05100	0.3201	-0.5905	0.6925	0.16	0.8740
		β_1	-0.00329	0.007839	-0.01869	0.01211	-0.42	0.6745
thorny skate	0.5065	β_o	0.1473	0.1817	-0.2144	0.5089	0.81	0.4201
		β_1	-0.00248	0.002861	-0.00810	0.003134	-0.87	0.3861
witch flounder	0.2964	β_o	-0.2420	0.1569	-0.5563	0.07234	-1.54	0.1287
		β_1	0.004826	0.004350	-0.00371	0.01336	1.11	0.2676
yellowtail flounder	0.3760	β_o	0.1492	0.1831	-0.2296	0.5280	0.81	0.4235
		β_1	-0.00542	0.003324	-0.01195	0.001115	-1.63	0.1039

Table 12. Comparison of fit statistics for the ME2 AR and random intercept (RI) model results. RPL: $-2\times$ Restricted Log Pseudo-Likelihood. GCS: Generalized Chi-Square.

	RPL		GCS	
Species	ME2 AR	ME2 RI	ME2 AR	ME2 RI
American plaice	7069.76	7220.47	2025.67	2444.13
Atlantic Cod	4419.81	4696.91	1022.01	1597.35
deep water redfish	3347.25	3842.28	1002.28	1814.92
Greenland halibut	2195.83	2196.35	575.88	588.89
thorny skate	4018.58	4110.75	891.09	1174.73
witch flounder	3228.75	3250.55	979.97	1071.17
yellowtail flounder	1471.86	1519.19	517.94	656.86

Table 13. Scanmar summary statistics for the Wilfred Templeman (WT) and the Alfred Needler (AN). N is the number of hauls. Std. Dev. is the standard deviation. CV is the coefficient of variation. set CV is the average within-set CV.

Survey	Statistic	Doors		Wings		Opening	
		WT	AN	WT	AN	WT	AN
Spring	N	45	45	63	63	61	61
	Mean	48.58	46.99	17.06	15.81	3.80	4.50
	Std. Dev.	5.61	6.06	1.74	1.14	0.41	0.48
	CV	11.55	12.89	10.20	7.24	10.78	10.64
	Minimum	32.12	35.78	13.99	11.37	3.09	3.74
	Median	48.26	47.05	16.84	15.87	3.76	4.53
	Maximum	57.22	62.61	21.03	17.72	5.12	5.75
	set CV	5.44	11.00	2.62	1.54	0.68	0.84
Fall	N	39	39	38	38	38	38
	Mean	44.06	45.83	16.28	15.74	5.03	4.38
	Std. Dev.	4.84	5.87	1.21	1.28	0.69	0.78
	CV	10.98	12.82	7.40	8.15	13.66	17.80
	Minimum	35.76	34.85	13.91	12.10	4.25	1.19
	Median	46.21	47.66	16.67	16.27	4.80	4.27
	Maximum	55.03	63.75	18.95	17.21	7.19	5.99
	set CV	3.16	4.63	1.61	2.08	1.84	1.19

Table 14. FE2 model results with swept area adjustments. Se - standard error. L,U - intervals. pv is the profile likelihood confidence intervals. pv is the χ^2 p-value. Significant estimates in bold.

Species	$\hat{\phi}$		Estimate	Se	95% L	95% U	χ^2	pv
American plaice	1.37	β_o	**0.2013**	0.0691	0.0659	0.3369	8.49	0.0036
		β_1	**-0.0094**	0.0026	-0.0145	-0.0043	12.85	0.0003
Atlantic Cod	1.76	β_o	**0.5430**	0.1498	0.2504	0.8381	13.14	0.0003
		β_1	**-0.0098**	0.0029	-0.0156	-0.0041	11.09	0.0009
deep water redfish	2.01	β_o	**0.2468**	0.1047	0.0415	0.4520	5.55	0.0184
		β_1	**-0.0229**	0.0045	-0.0318	-0.0141	25.80	$<.0001$
Greenland halibut	1.18	β_o	-0.1368	0.3305	-0.7872	0.5102	0.17	0.6790
		β_1	-0.0009	0.0083	-0.0171	0.0154	0.01	0.9151
thorny skate	1.37	β_o	0.0166	0.1211	-0.2208	0.2543	0.02	0.8912
		β_1	-0.0051	0.0027	-0.0103	0.0002	3.52	0.0605
witch flounder	1.26	β_o	-0.1811	0.1231	-0.4228	0.0597	2.17	0.1412
		β_1	-0.0024	0.0042	-0.0107	0.0059	0.32	0.5743
yellowtail flounder	1.38	β_o	0.1320	0.1238	-0.1109	0.3746	1.14	0.2865
		β_1	-0.0073	0.0042	-0.0155	0.0009	3.04	0.0812

Table 15. ME2 AR model results with swept area adjustments. Se - standard error. L,U - Wald confidence intervals. pv is the t-statistic p-value. Significant estimates in bold.

Species	$\hat{\sigma}^2/\hat{\gamma}$		Estimate	Se	95% L	95% U	t	pv
American plaice	0.6361	β_o	0.2434	0.1513	-0.05669	0.5434	1.61	0.1108
	0.9580	β_1	-0.00764	0.004404	-0.01628	0.000996	-1.73	0.0829
Atlantic Cod	1.6343	β_o	-0.2178	0.2703	-0.7547	0.3191	-0.81	0.4225
	0.9741	β_1	0.001189	0.004914	-0.00845	0.01083	0.24	0.8089
deep water redfish	1.0628	β_o	0.1730	0.2279	-0.2826	0.6286	0.76	0.4507
	0.9315	β_1	**-0.01807**	0.008602	-0.03495	-0.00119	-2.10	0.0359
Greenland halibut	0.1392	β_o	0.06855	0.3490	-0.6308	0.7679	0.20	0.8450
	0.9422	β_1	-0.00562	0.008636	-0.02258	0.01135	-0.65	0.5157
thorny skate	1.0051	β_o	-0.01990	0.2613	-0.5401	0.5003	-0.08	0.9395
	0.9106	β_1	-0.00145	0.004774	-0.01082	0.007920	-0.30	0.7615
witch flounder	0.3707	β_o	-0.2063	0.1957	-0.5984	0.1857	-1.05	0.2963
	0.9630	β_1	0.002194	0.006057	-0.00969	0.01408	0.36	0.7173
yellowtail flounder	0.4882	β_o	0.1295	0.2711	-0.4313	0.6903	0.48	0.6373
	0.9633	β_1	-0.00685	0.007903	-0.02238	0.008672	-0.87	0.3862

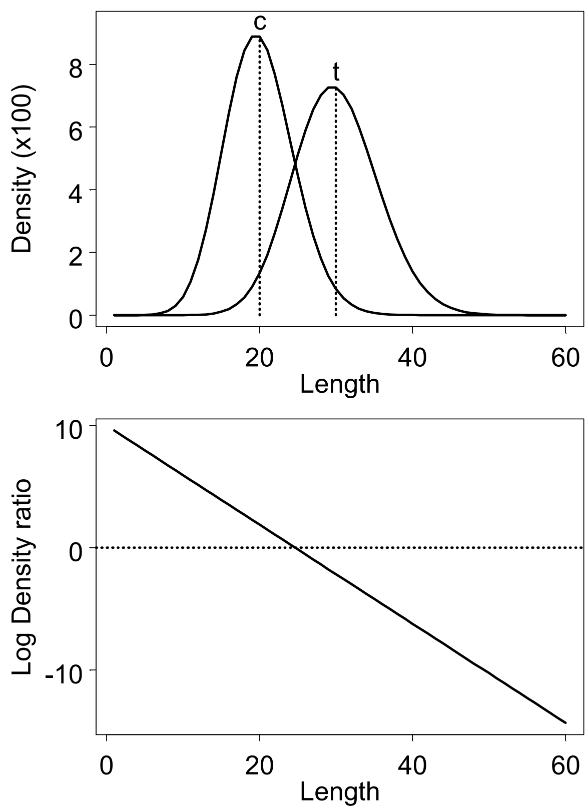

Figure 1. Top panel: Hypothetical length distributions sampled by each trawl, λ_{lc} and λ_{lt} Bottom panel: log density ratio, $\delta_l = \log(\lambda_{lc}/\lambda_{lt})$.

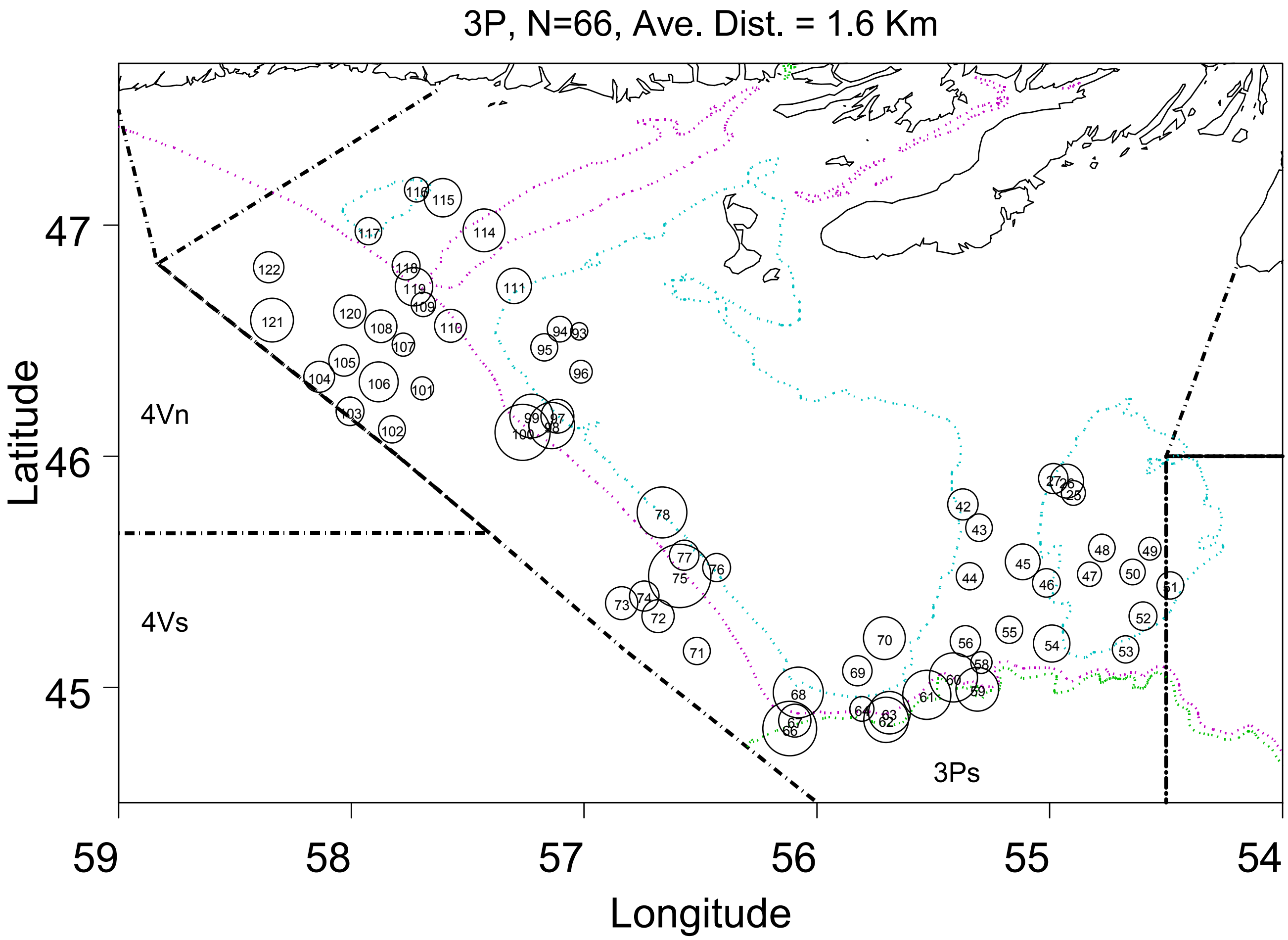

Figure 2. Locations of paired fishing sets (i.e. average latitude and longitude) in NAFO Subdivision 3Ps. Set numbers are shown. The bubble area is proportional to the distance between paired tows. The number of paired sets (N) and the average distance between tows (Ave. Dist.) are listed at the top. 100 m (blue), 300 m (purple), and 500 m (green) depth contours are shown.

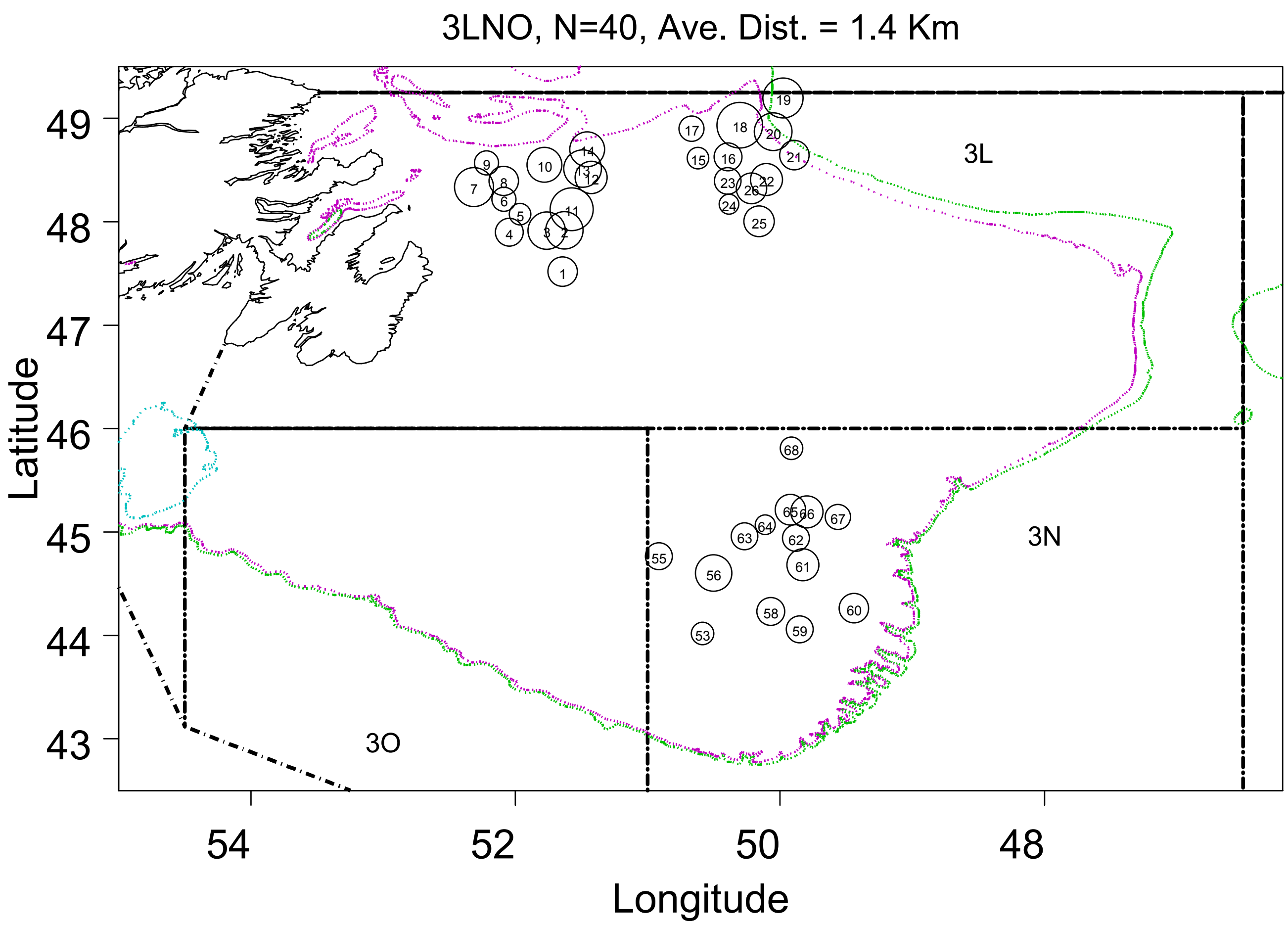

Figure 3. Locations of paired fishing sets (i.e. average latitude and longitude) in NAFO Divisions 3LNO. Set numbers are shown. The bubble area is proportional to the distance between paired tows. The number of paired sets (N) and the average distance between tows (Ave. Dist.) are listed at the top. 100 m (blue), 300 m (purple), and 500 m (green) depth contours are shown.

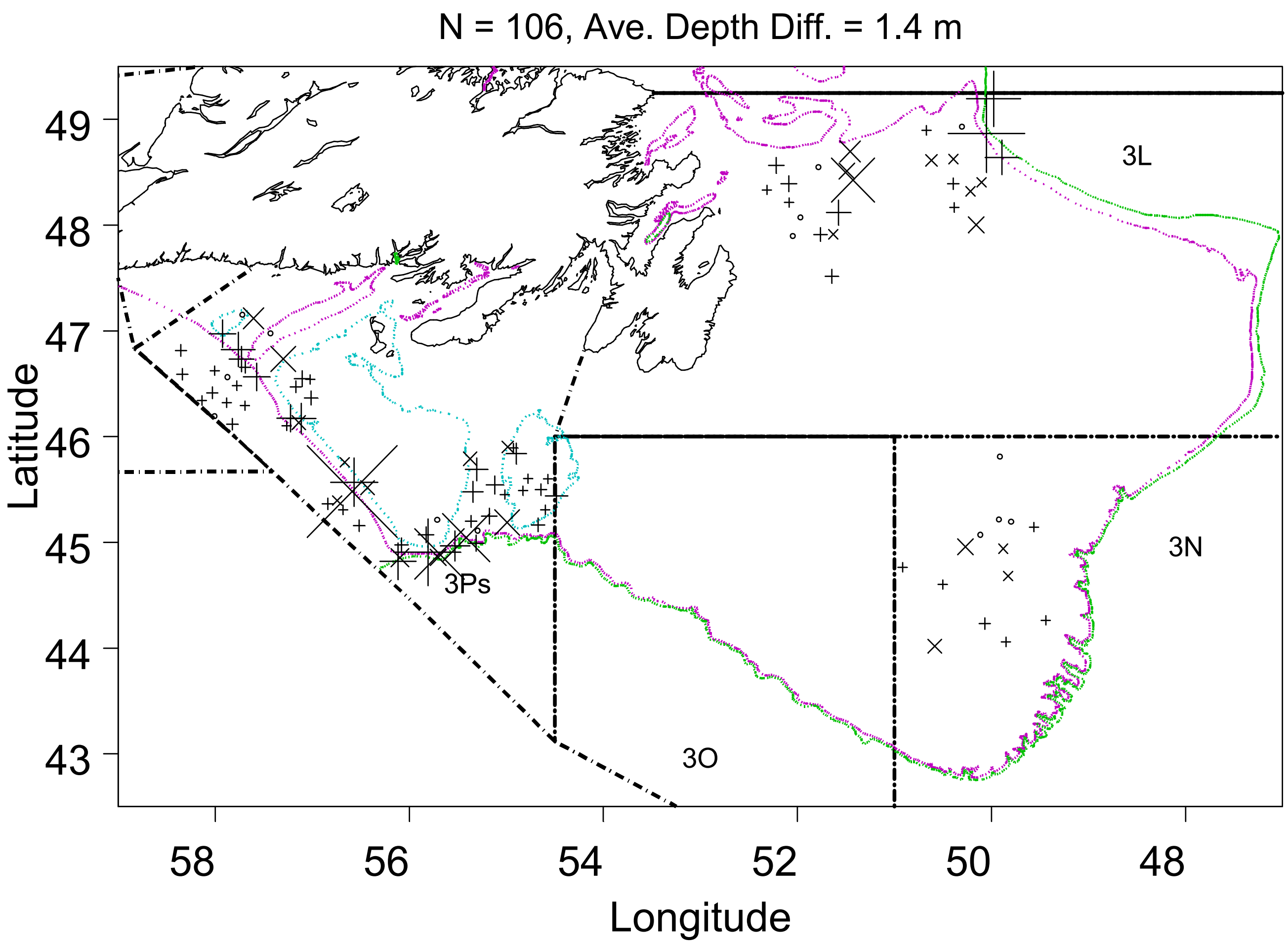

Figure 4. Difference between tow depths (WT minus AN). Plotting symbols are proportional to the difference, $\times$ is negative, $+$ is positive, $\circ$ is zero. The number of paired sets (N) and the average depth difference (Ave. Depth Diff.) are listed at the top. 100 m (blue), 300 m (purple), and 500 m (green) depth contours are shown.

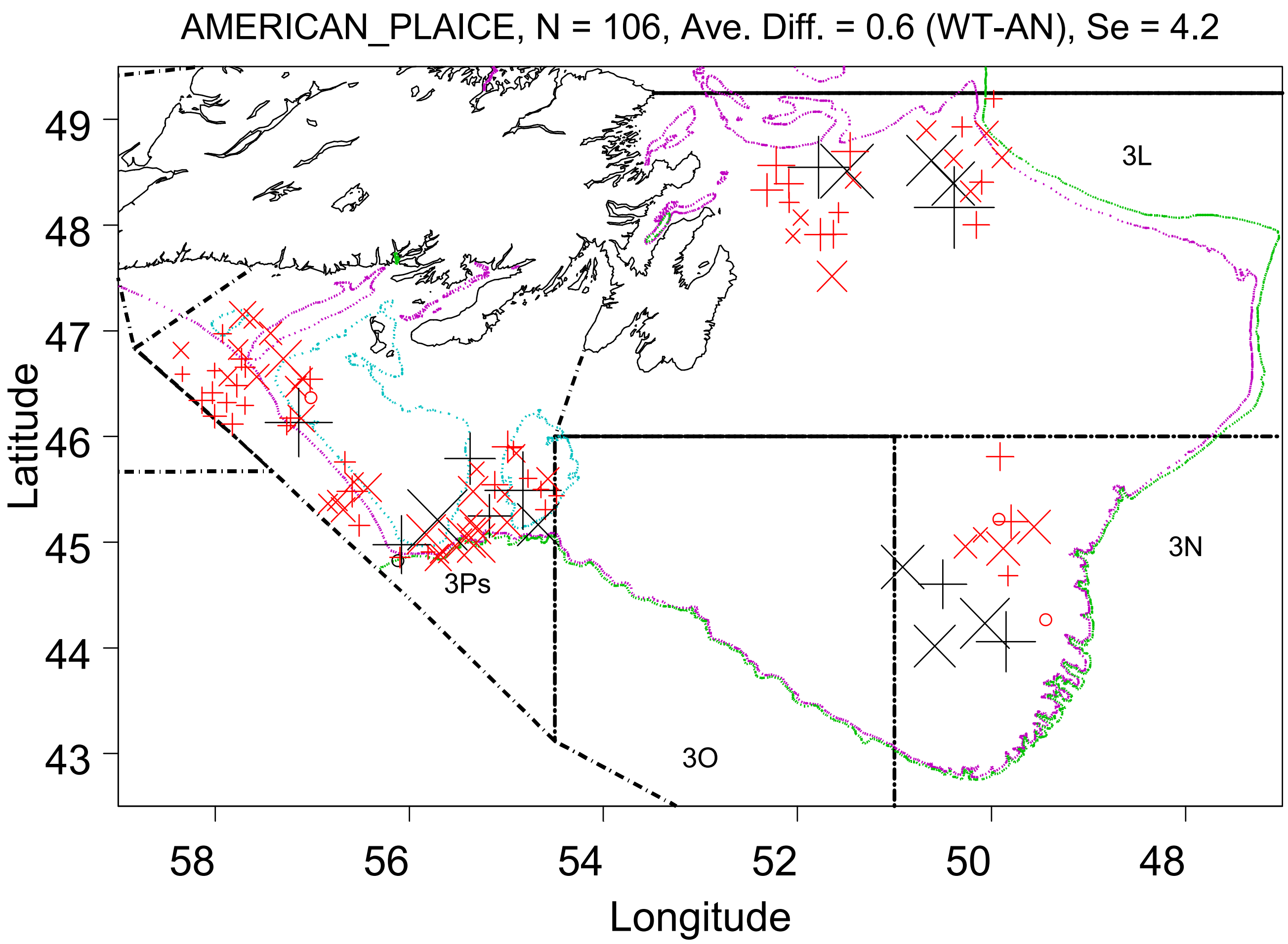

Figure 5. Differences between American plaice catches (WT minus AN), divided by their Poisson standard error. Catches are scaled for subsampling, and standardized for tow distance. Plotting symbols are proportional to the difference: × - negative, + - positive, ○ - zero, ○ - no catch in either pair. Black ×, + are potential outliers. The number of paired sets (N), the average catch difference (Ave. Diff.), and the standard error of the differences (Se) are listed at the top. 100 m (blue), 300 m (purple), and 500 m (green) depth contours are shown.

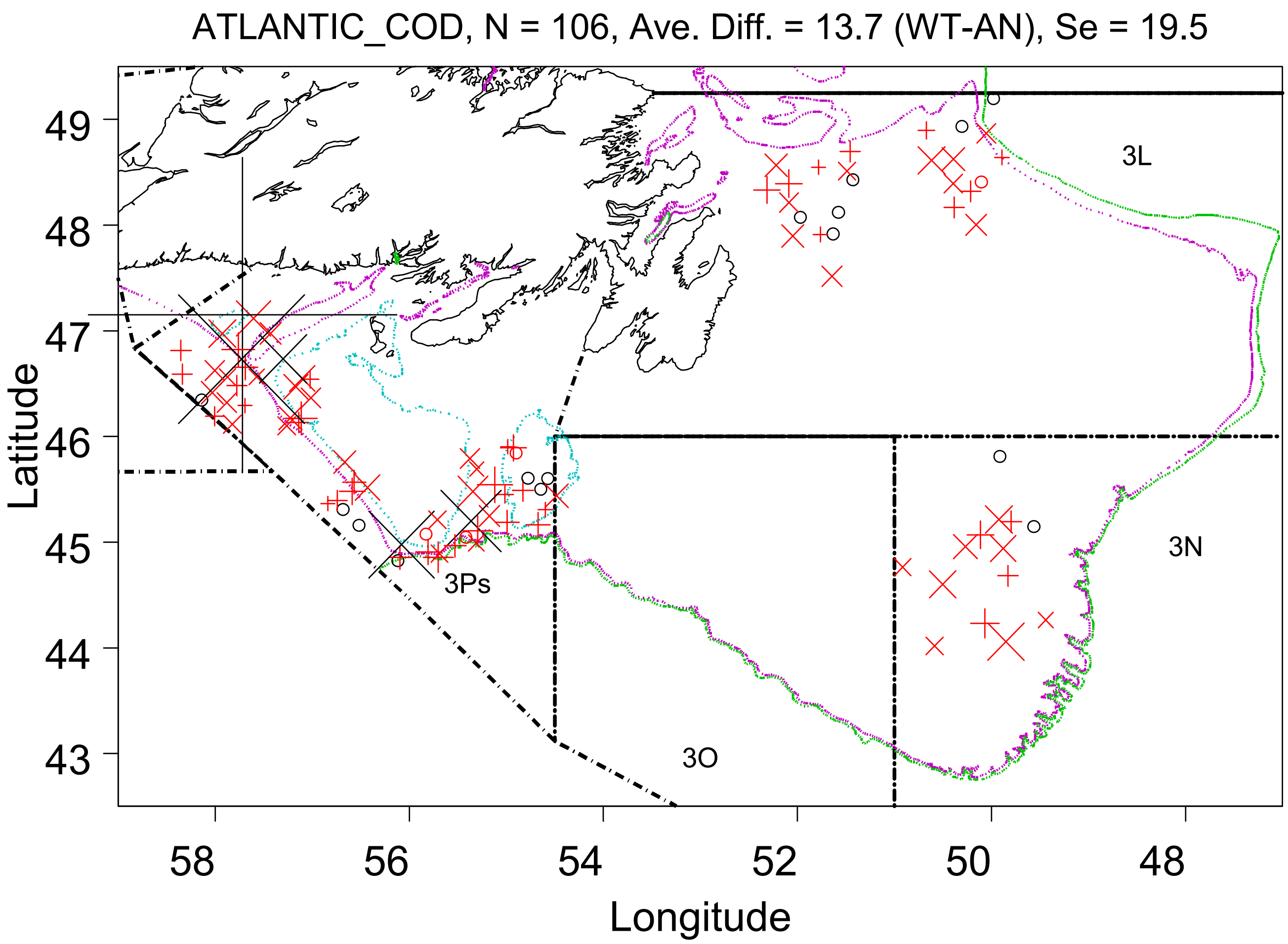

Figure 6. Differences between Atlantic cod catches (WT minus AN), divided by their Poisson standard error. Catches are scaled for subsampling, and standardized for tow distance. Plotting symbols are proportional to the difference: × - negative, + - positive, ∘ - zero, ∘ - no catch in either pair. Black ×, + are potential outliers. The number of paired sets (N), the average catch difference (Ave. Diff.), and the standard error of the differences (Se) are listed at the top. 100 m (blue), 300 m (purple), and 500 m (green) depth contours are shown.

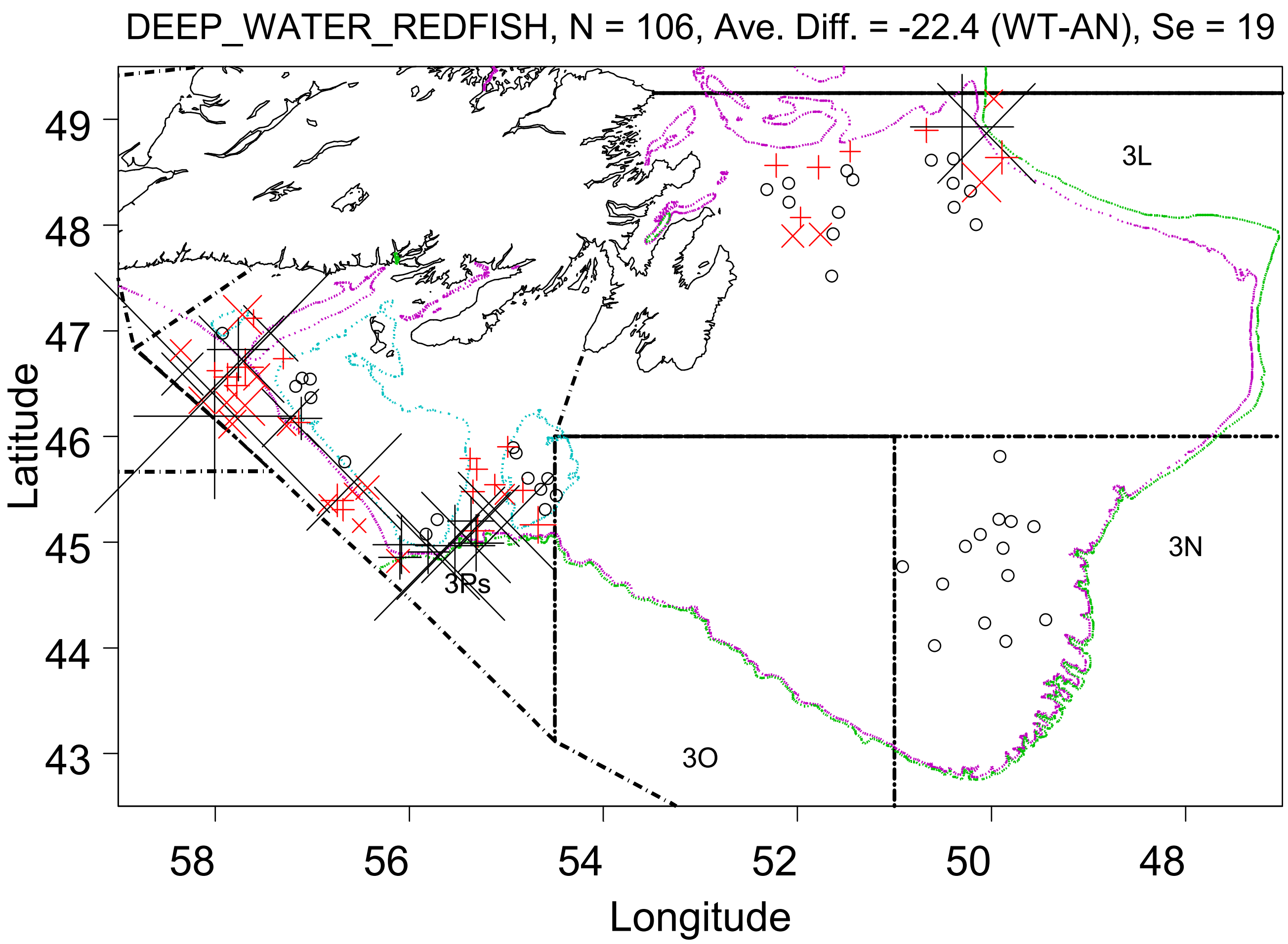

Figure 7. Differences between deepwater redfish catches (WT minus AN), divided by their Poisson standard error. Catches are scaled for subsampling, and standardized for tow distance. Plotting symbols are proportional to the difference: × - negative, + - positive, ○ - zero, ○ - no catch in either pair. Black ×, + are potential outliers. The number of paired sets (N), the average catch difference (Ave. Diff.), and the standard error of the differences (Se) are listed at the top. 100 m (blue), 300 m (purple), and 500 m (green) depth contours are shown.

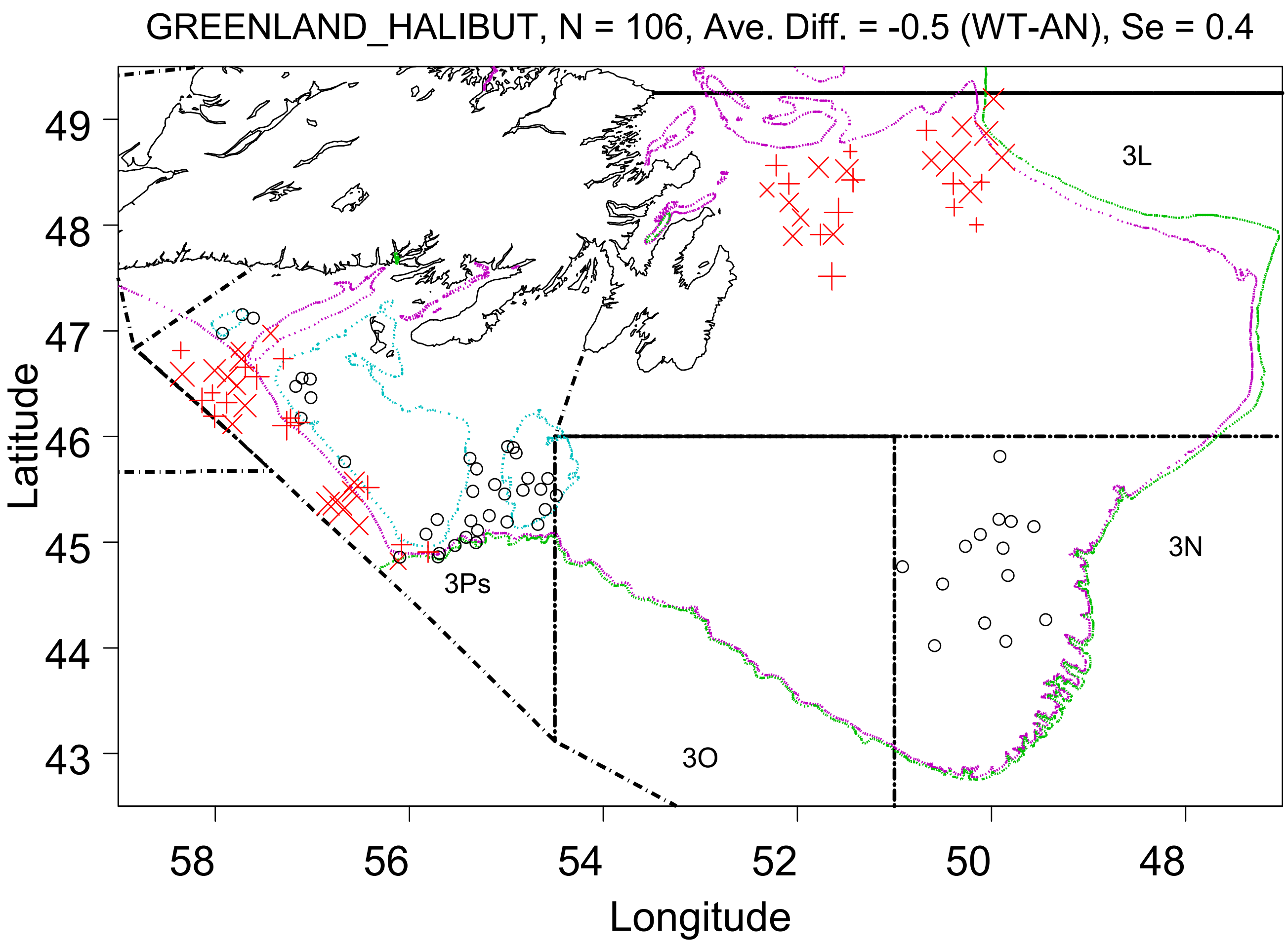

Figure 8. Differences between Greenland halibut catches (WT minus AN), divided by their Poisson standard error. Catches are scaled for subsampling, and standardized for tow distance. Plotting symbols are proportional to the difference: × - negative, + - positive, ○ - zero, ○ - no catch in either pair. Black ×, + are potential outliers. The number of paired sets (N), the average catch difference (Ave. Diff.), and the standard error of the differences (Se) are listed at the top. 100 m (blue), 300 m (purple), and 500 m (green) depth contours are shown.

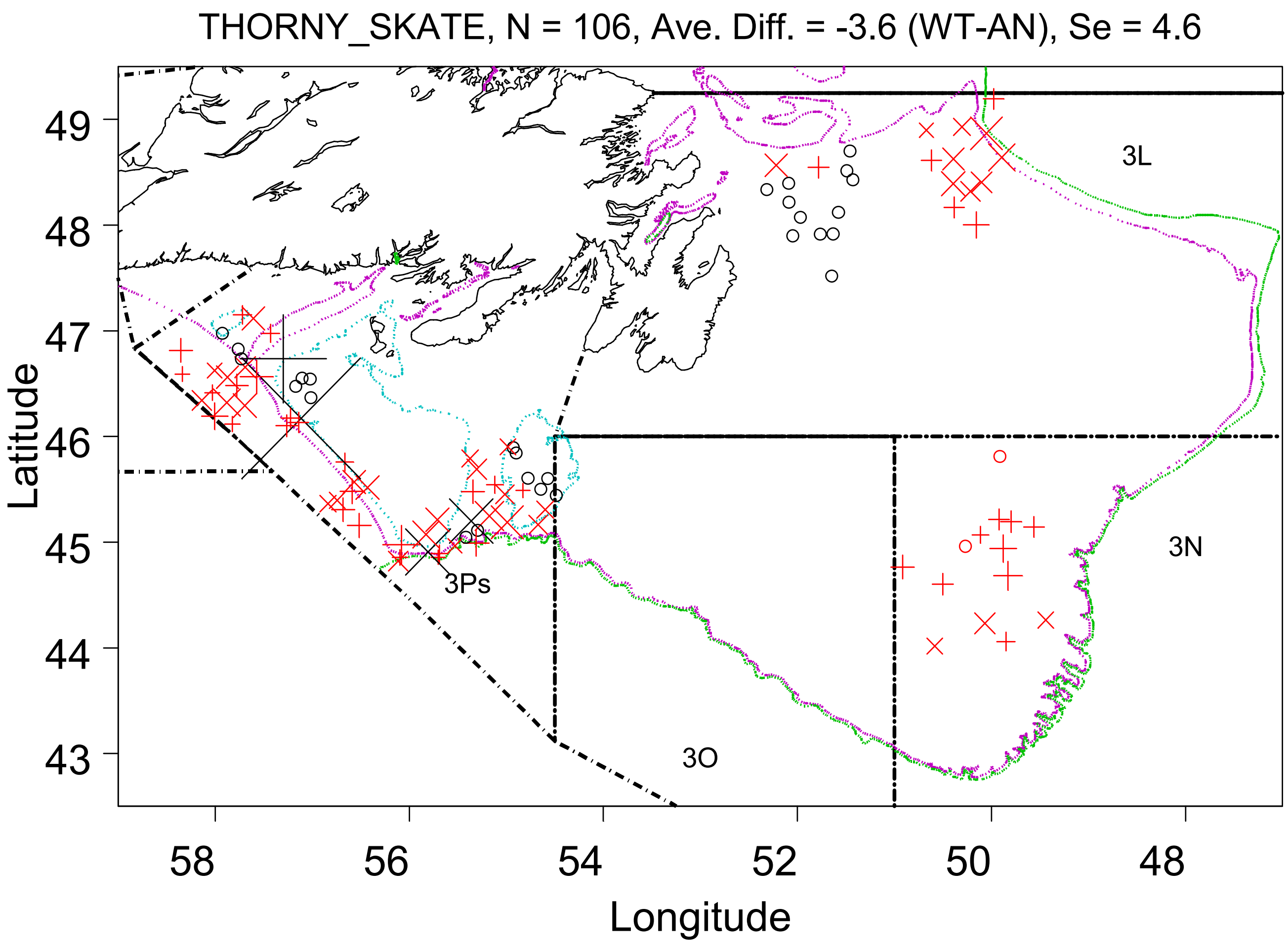

Figure 9. Differences between thorny skate catches (WT minus AN), divided by their Poisson standard error. Catches are scaled for subsampling, and standardized for tow distance. Plotting symbols are proportional to the difference: × - negative, + - positive, ○ - zero, ○ - no catch in either pair. Black ×, + are potential outliers. The number of paired sets (N), the average catch difference (Ave. Diff.), and the standard error of the differences (Se) are listed at the top. 100 m (blue), 300 m (purple), and 500 m (green) depth contours are shown.

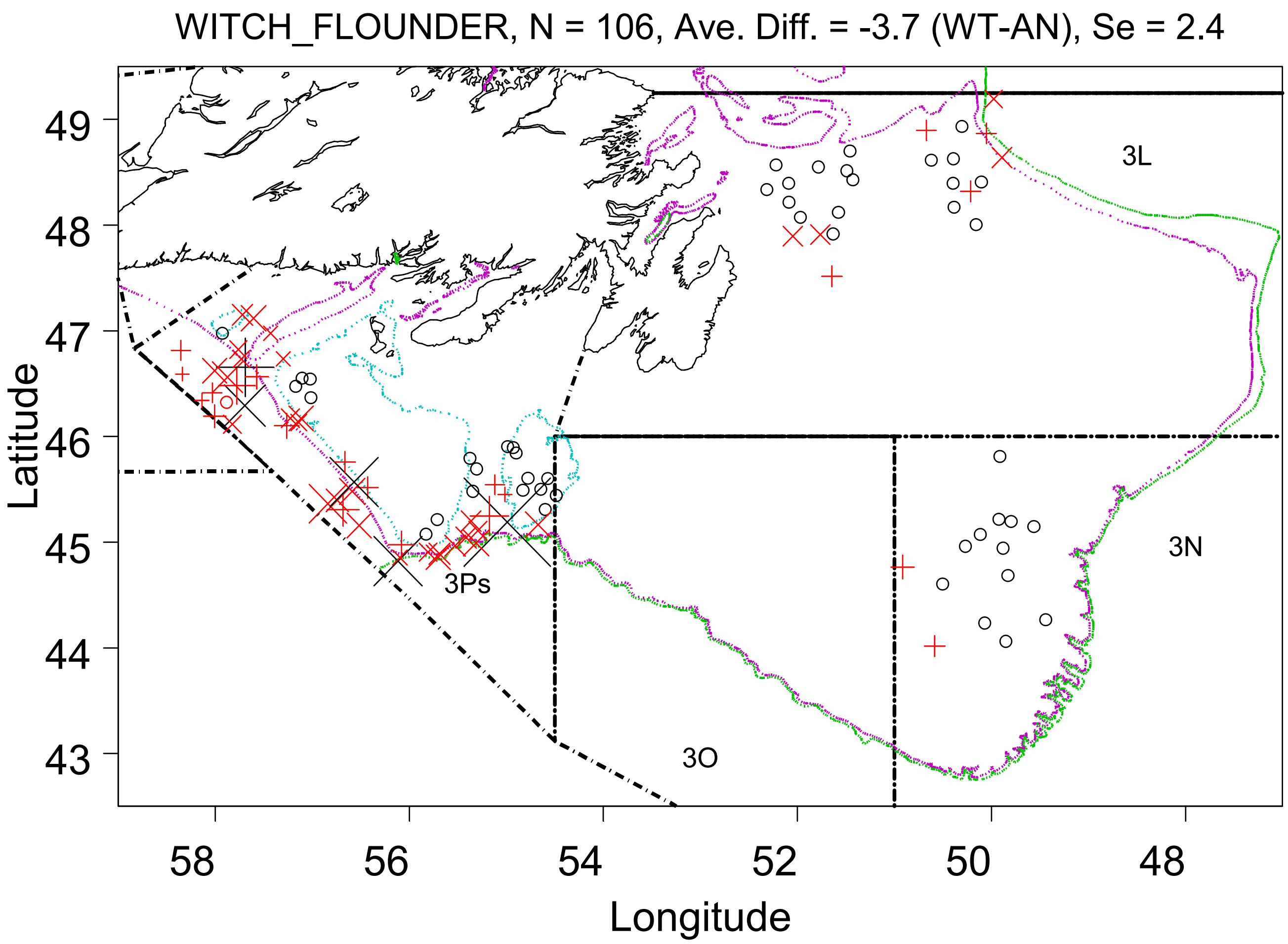

Figure 10. Differences between witch flounder catches (WT minus AN), divided by their Poisson standard error. Catches are scaled for subsampling, and standardized for tow distance. Plotting symbols are proportional to the difference: × - negative, + - positive, ○ - zero, ○ - no catch in either pair. Black ×, + are potential outliers. The number of paired sets (N), the average catch difference (Ave. Diff.), and the standard error of the differences (Se) are listed at the top. 100 m (blue), 300 m (purple), and 500 m (green) depth contours are shown.

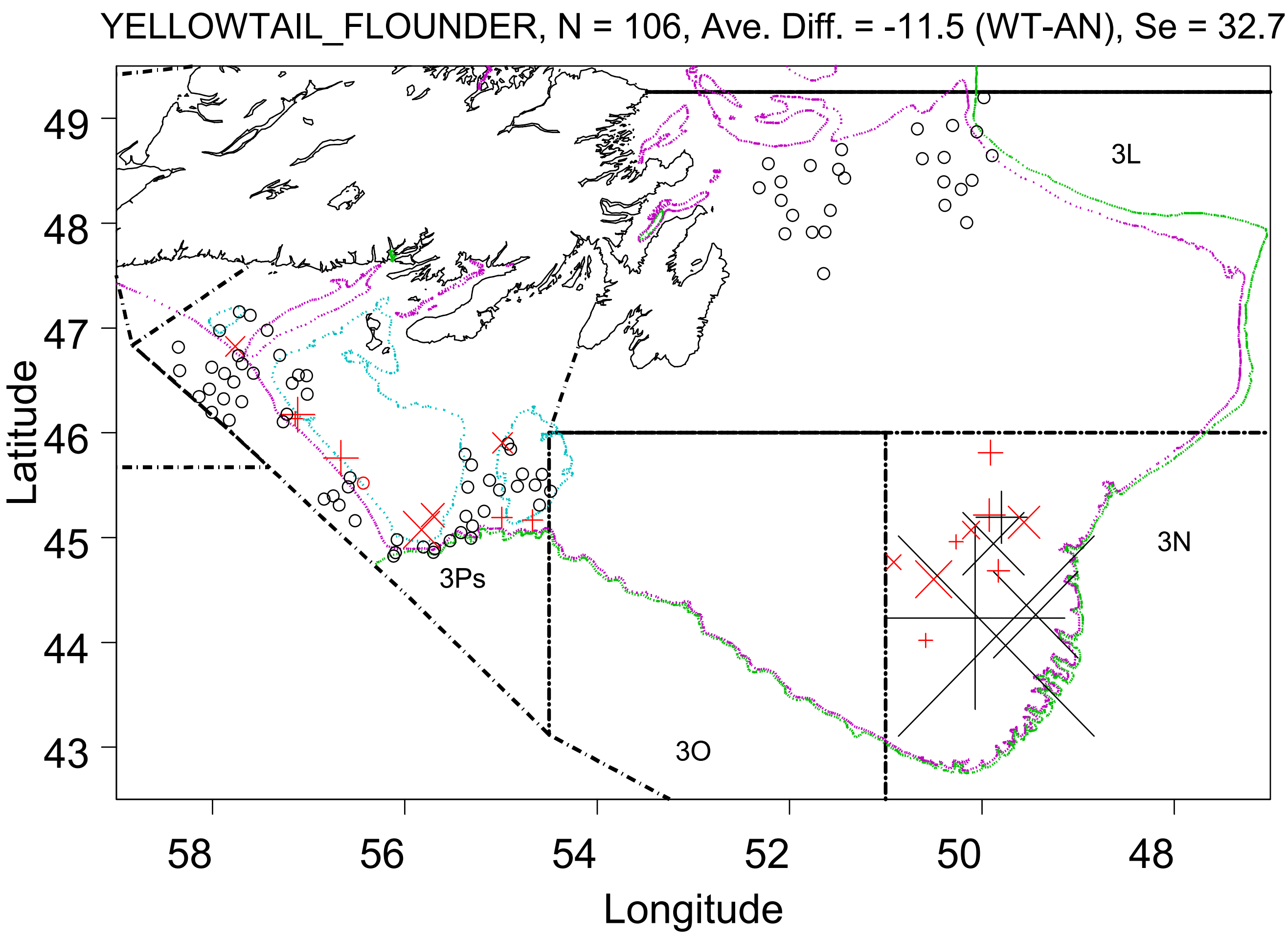

Figure 11. Differences between yellowtail flounder catches (WT minus AN), divided by their Poisson standard error. Catches are scaled for subsampling, and standardized for tow distance. Plotting symbols are proportional to the difference: × - negative, + - positive, ○ - zero, ○ - no catch in either pair. Black ×, + are potential outliers. The number of paired sets (N), the average catch difference (Ave. Diff.), and the standard error of the differences (Se) are listed at the top. 100 m (blue), 300 m (purple), and 500 m (green) depth contours are shown.

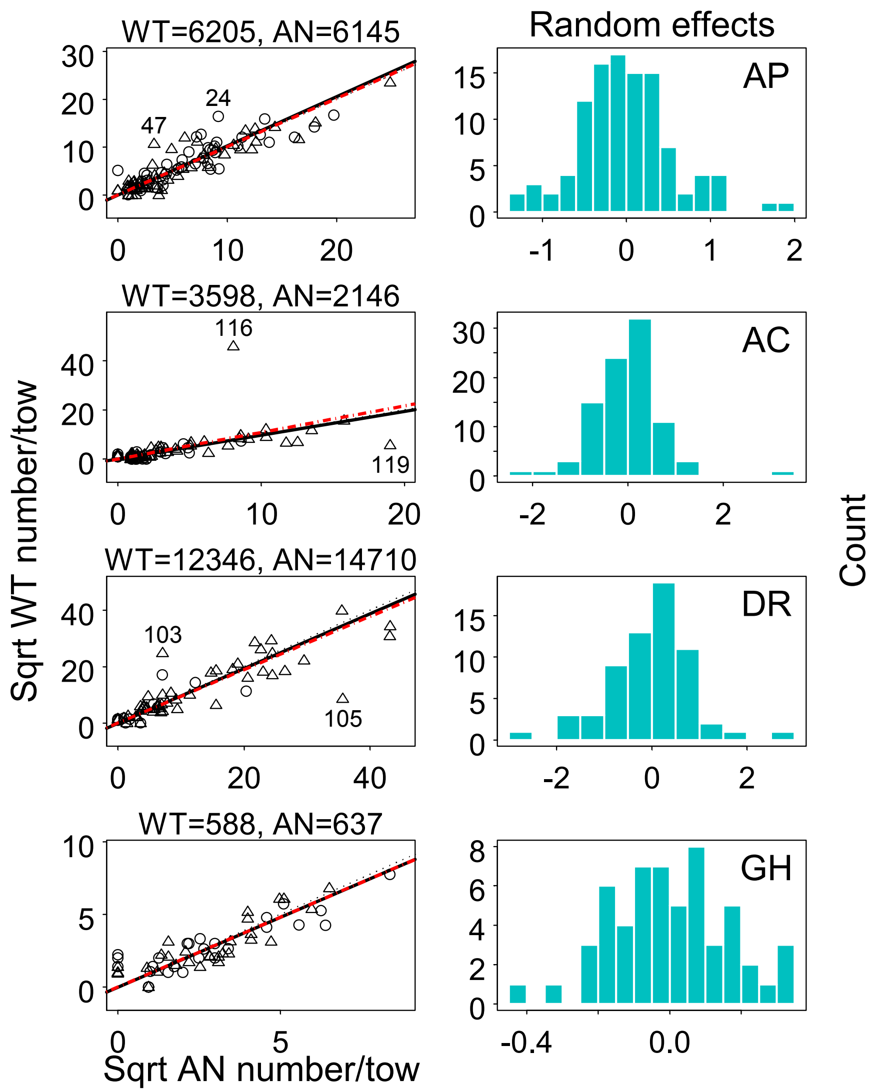

Figure 12. Left column: Total scaled catches (r^*) from each haul, AN vs WT. Plotting symbols: ○ - Fall survey; △ - spring survey. Totals for all sets are listed at the top. The dotted line has a slope of one. The red dashed line has a slope equal to the estimated relative efficiency (ρ) from the FEP1 model. Sets with large deviance residuals are indicated with their set number. The solid line represents the mean ρ estimated from the MEP1 model. Right column: Predicted random effects histograms. Rows are for species, with codes indicated in the upper right-hand corner of the histograms: AP - American plaice; AC - Atlantic cod; DR - deepwater redfish; GH - Greenland halibut.

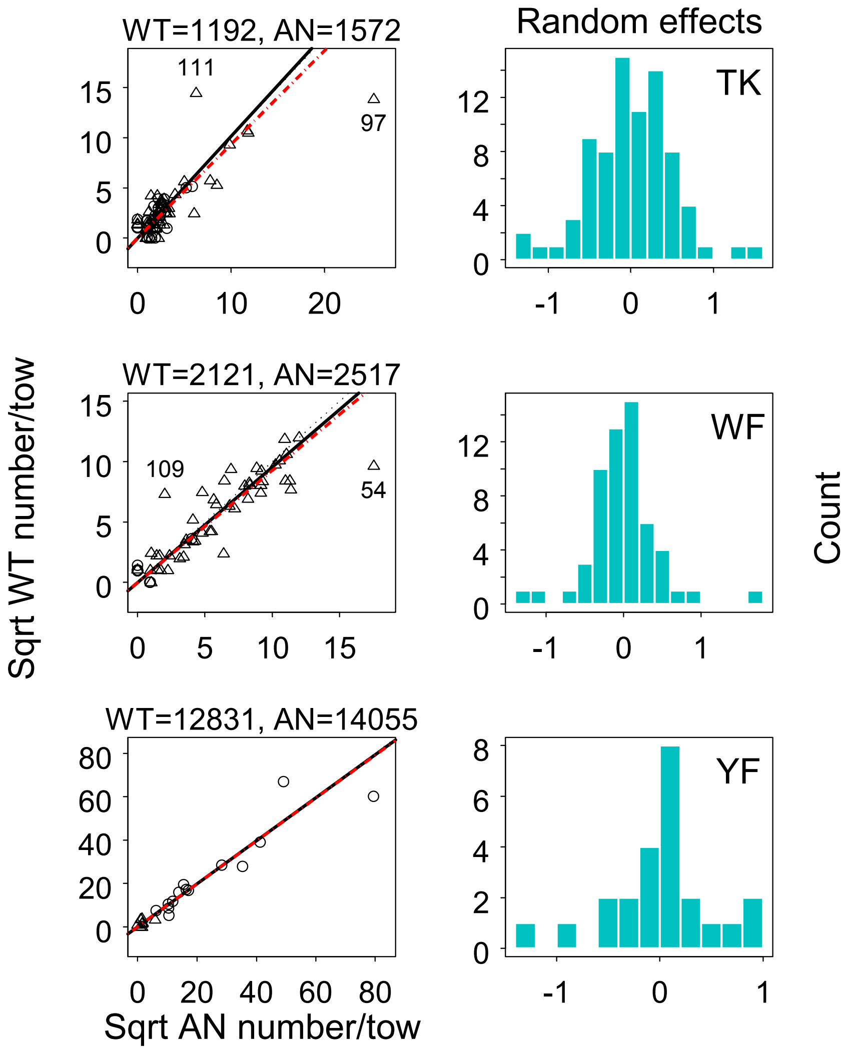

Figure 13. Left column: Total scaled catches (r^*) from each haul, AN vs WT. Plotting symbols: ○ - Fall survey; △ - spring survey. Totals for all sets are listed at the top. The dotted line has a slope of one. The red dashed line has a slope equal to the estimated relative efficiency (ρ) from the FEP1 model. Sets with large deviance residuals are indicated with their set number. The solid line represents the mean ρ estimated from the MEP1 model. Right column: Predicted random effects histograms. Rows are for species, with codes indicated in the upper right-hand corner of the histograms: TK - thorny skate; WF - witch flounder; YF - yellowtail flounder.

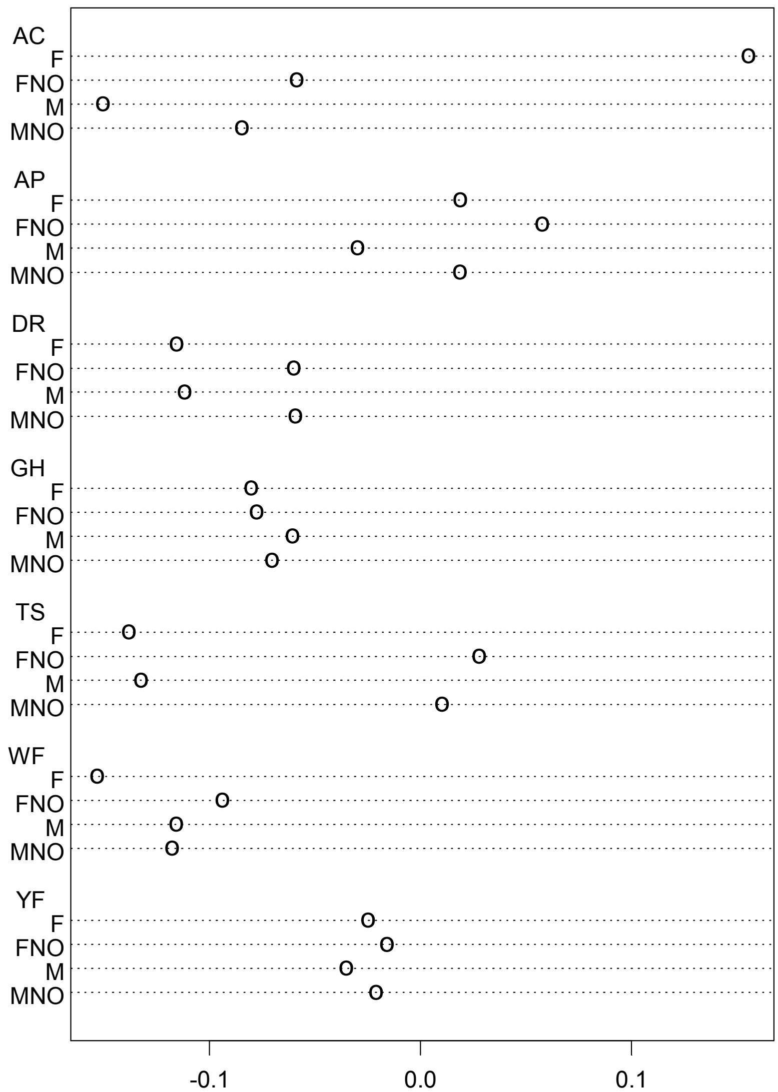

Figure 14. Estimates of $\hat{\beta}_o$ from the FEP1 (F) and MEP1 (M) models, and models with 2 potential outliers removed (FNO, MNO). Species codes: AP - American plaice; AC - Atlantic cod; DR - deepwater redfish; GH - Greenland halibut; TK - thorny skate; WF - witch flounder; YF - yellowtail flounder.

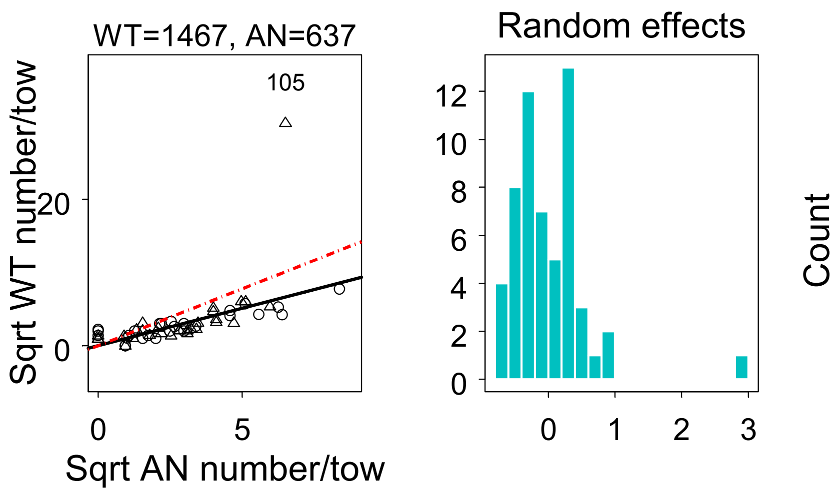

Figure 15. Outlier added for set 105. Left column: Total scaled catches (r^*) from each haul for Greenland halibut, AN vs WT. Plotting symbols: ○ - Fall survey; △ - spring survey. Totals for all sets are listed at the top. The dotted line has a slope of one. The red dashed line has a slope equal to the estimated relative efficiency (ρ) from the FEP1 model. Sets with large deviance residuals are indicated with their set number. The solid line represents the mean ρ estimated from the MEP1 model. Right column: Predicted random effects histograms.

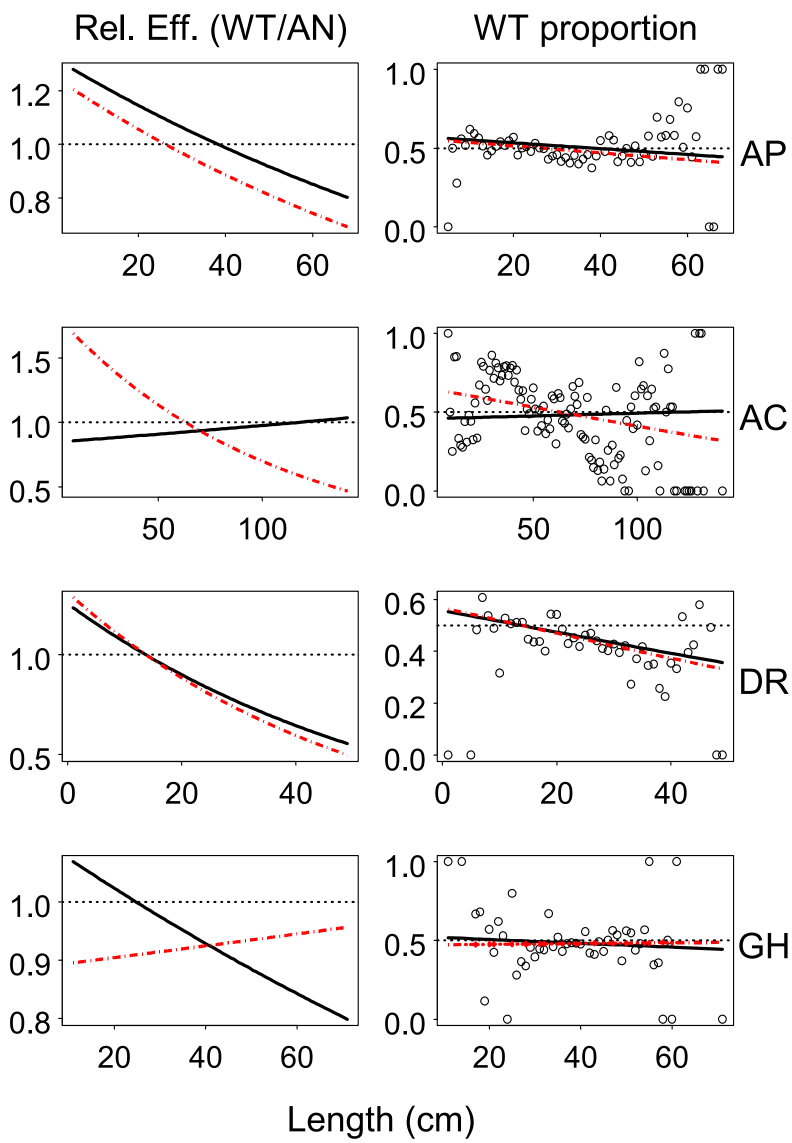

Figure 16. Left column: ME2 model estimated relative efficiency ($\hat{\rho}_l$; solid line) compared to the FE2 model estimates (red dashed line). Right column: Observed (○'s) and estimated (lines) proportions of WT scaled catches. Rows are for species, with codes indicated at the right-hand side: AP - American plaice; AC - Atlantic cod; DR - deepwater redfish; GH - Greenland halibut.

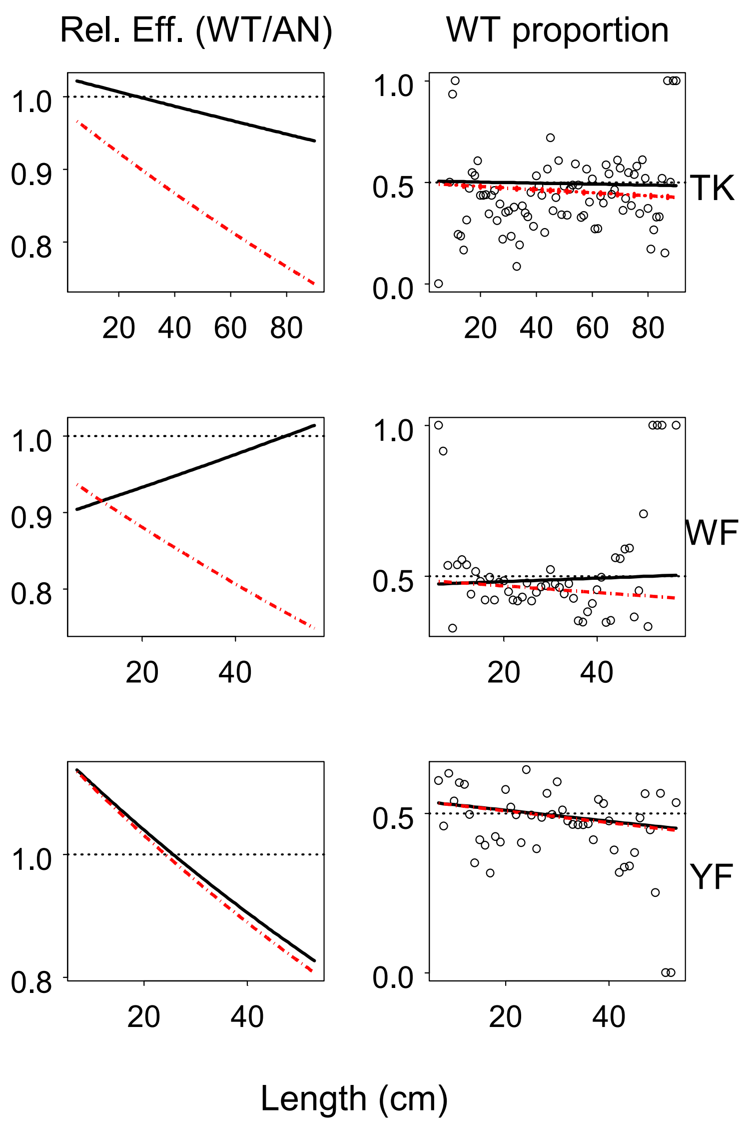

Figure 17. Left column: ME2 model estimated relative efficiency ($\hat{\rho}_l$; solid line) compared to the FE2 model estimates (red dashed line). Right column: Observed ($\circ$'s) and estimated (lines) proportions of WT scaled catches. Rows are for species, with codes indicated at the right-hand side: TK - thorny skate; WF - witch flounder; YF - yellowtail flounder.

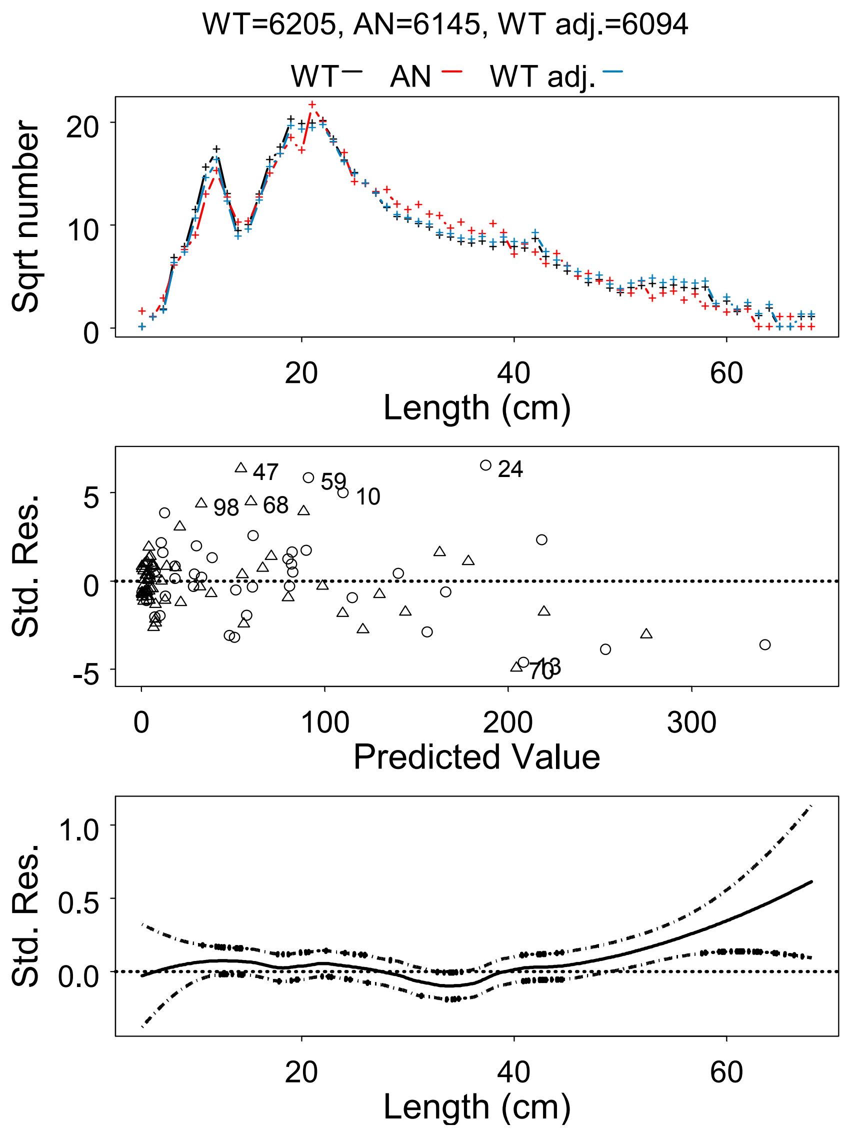

Figure 18. FE2 model results for American plaice. Total scaled catches (r^*) and WT catch adjusted by relative efficiency (ρ_l) are shown at the top. Top: Total length frequencies for all sets. Middle: Standardized (by st. err.) total chi-square residuals for each set, vs predicted WT catch. Plotting symbols: $\circ$ - Fall survey; $\triangle$ - spring survey. Residuals larger than ± 4 are identified by set number. Bottom: Standard chi-square residuals versus length. The solid line is a local linear smoother, and the dashed lines are 95% confidence intervals for the average residual.

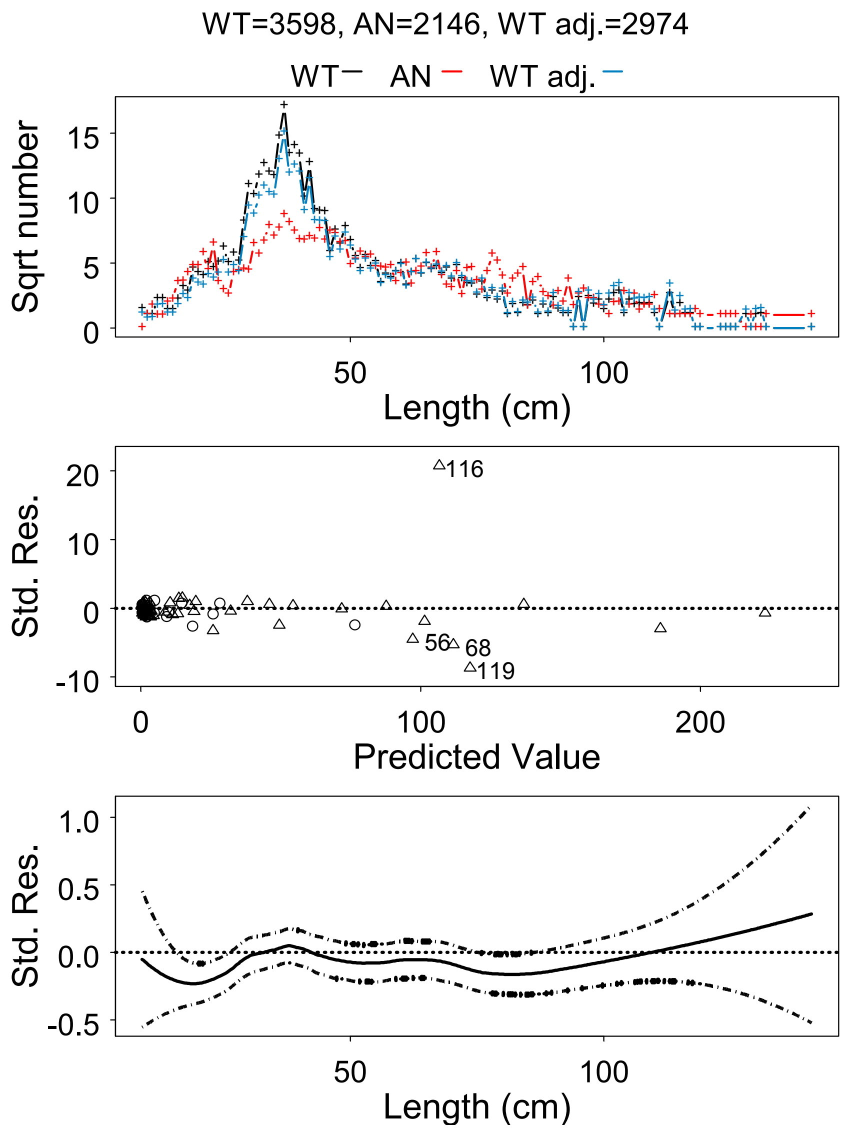

Figure 19. FE2 model results for Atlantic cod. Total scaled catches (r^*) and WT catch adjusted by relative efficiency (ρ_l) are shown at the top. Top: Total length frequencies for all sets. Middle: Standardized (by st. err.) total chi-square residuals for each set, vs predicted WT catch. Plotting symbols: ○ - Fall survey; △ - spring survey. Residuals larger than ± 4 are identified by set number. Bottom: Standard chi-square residuals versus length. The solid line is a local linear smoother, and the dashed lines are 95% confidence intervals for the average residual.

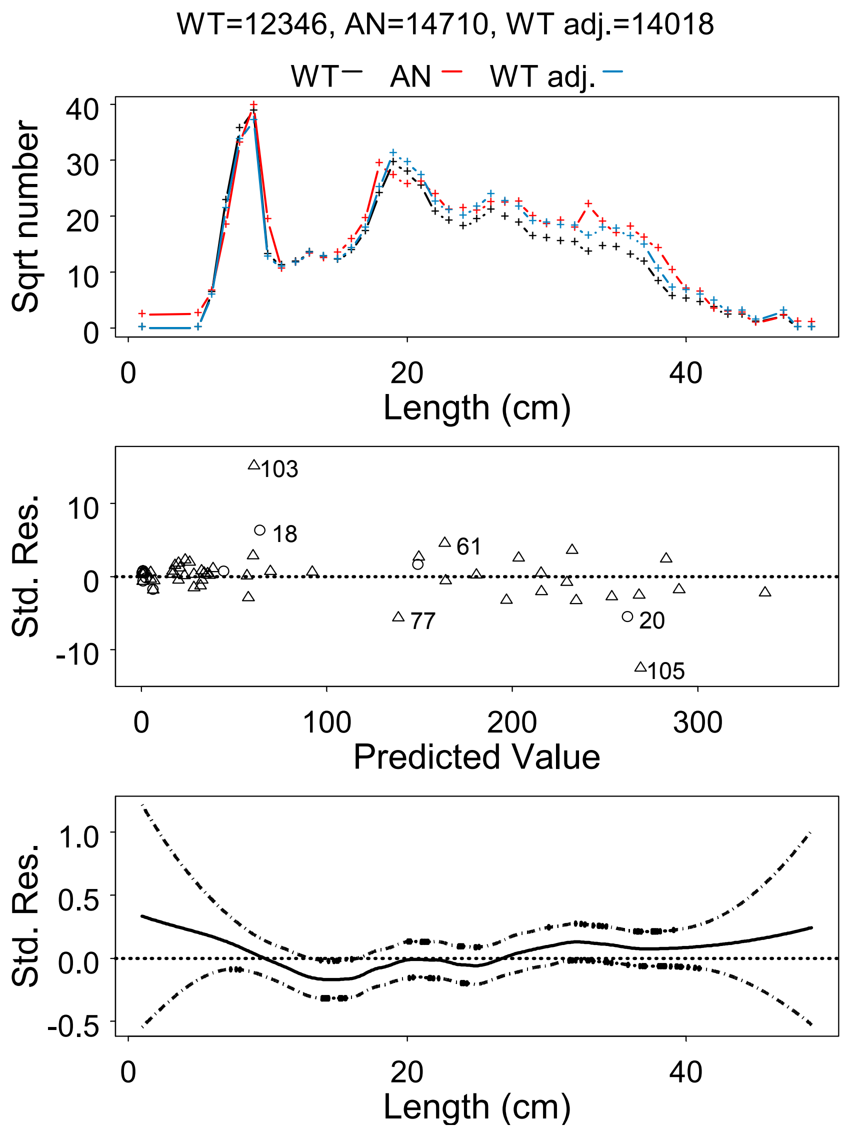

Figure 20. FE2 model results for deepwater redfish. Total scaled catches (r^*) and WT catch adjusted by relative efficiency (ρ_l) are shown at the top. Top: Total length frequencies for all sets. Middle: Standardized (by st. err.) total chi-square residuals for each set, vs predicted WT catch. Plotting symbols: $\circ$ - Fall survey; $\triangle$ - spring survey. Residuals larger than ± 4 are identified by set number. Bottom: Standard chi-square residuals versus length. The solid line is a local linear smoother, and the dashed lines are 95% confidence intervals for the average residual.

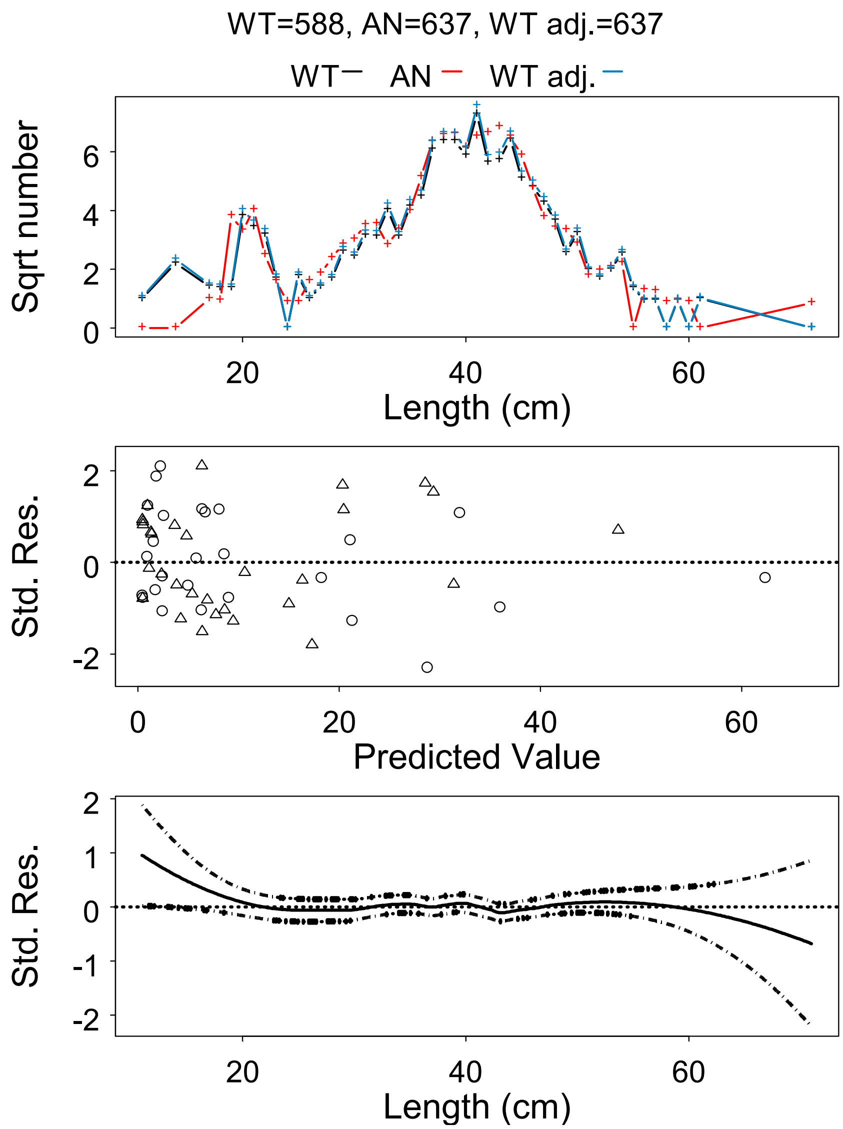

Figure 21. FE2 model results for Greenland halibut. Total scaled catches (r^*) and WT catch adjusted by relative efficiency (ρ_l) are shown at the top. Top: Total length frequencies for all sets. Middle: Standardized (by st. err.) total chi-square residuals for each set, vs predicted WT catch. Plotting symbols: $\circ$ - Fall survey; $\triangle$ - spring survey. Residuals larger than ± 4 are identified by set number. Bottom: Standard chi-square residuals versus length. The solid line is a local linear smoother, and the dashed lines are 95% confidence intervals for the average residual.

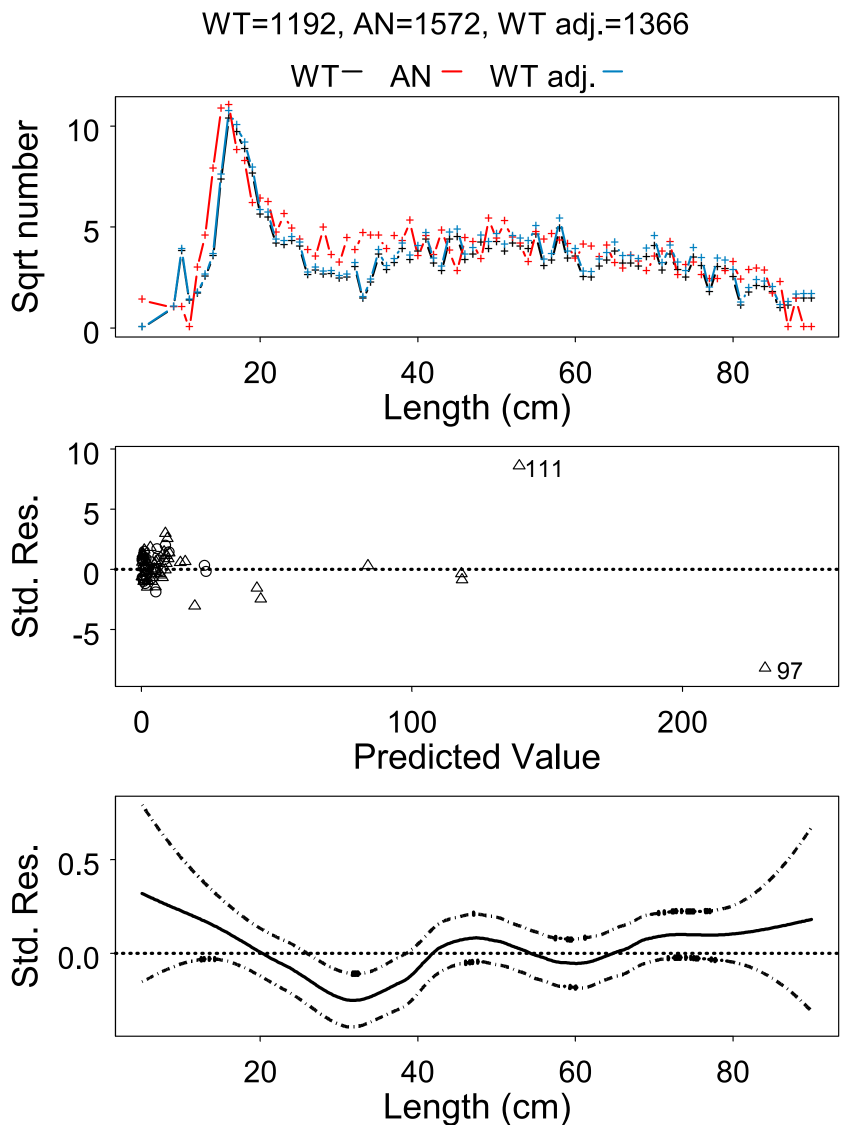

Figure 22. FE2 model results for thorny skate. Total scaled catches (r^*) and WT catch adjusted by relative efficiency (ρ_l) are shown at the top. Top: Total length frequencies for all sets. Middle: Standardized (by st. err.) total chi-square residuals for each set, vs predicted WT catch. Plotting symbols: $\circ$ - Fall survey; $\triangle$ - spring survey. Residuals larger than ± 4 are identified by set number. Bottom: Standard chi-square residuals versus length. The solid line is a local linear smoother, and the dashed lines are 95% confidence intervals for the average residual.

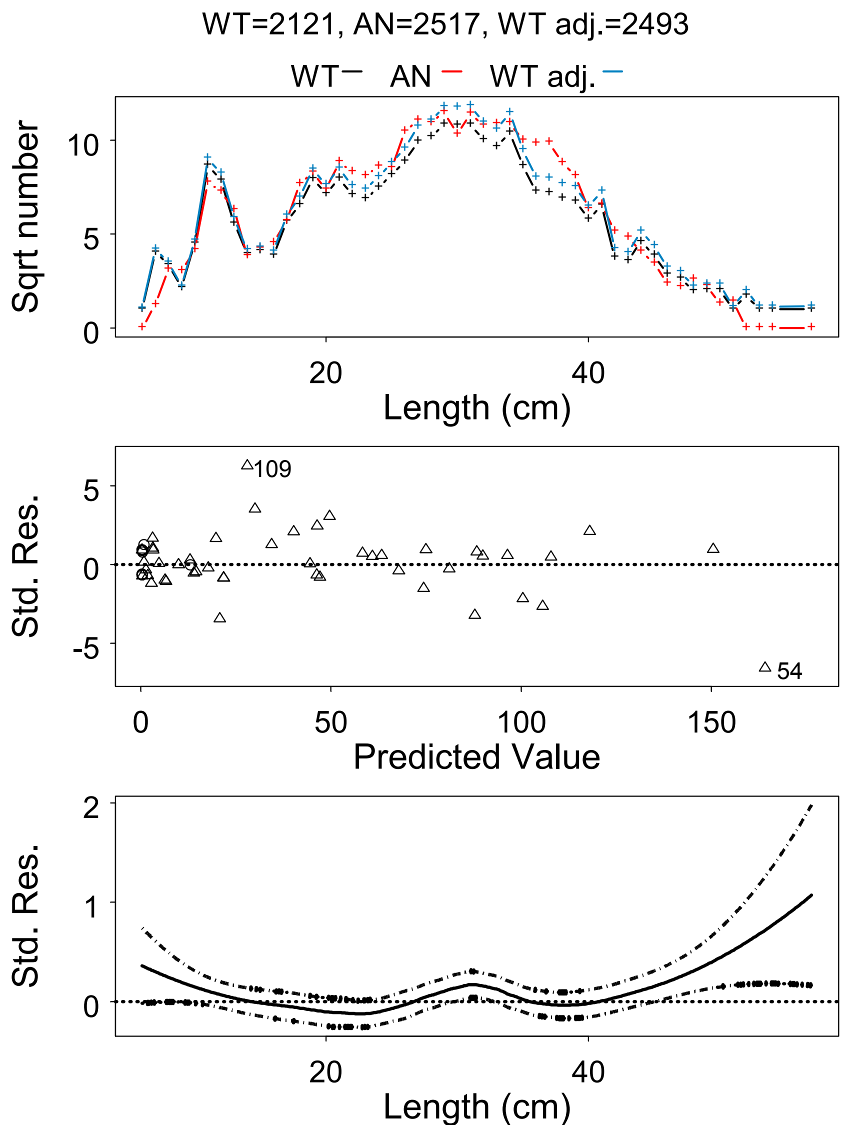

Figure 23. FE2 model results for witch flounder. Total scaled catches (r^*) and WT catch adjusted by relative efficiency (ρ_l) are shown at the top. Top: Total length frequencies for all sets. Middle: Standardized (by st. err.) total chi-square residuals for each set, vs predicted WT catch. Plotting symbols: $\circ$ - Fall survey; $\triangle$ - spring survey. Residuals larger than ± 4 are identified by set number. Bottom: Standard chi-square residuals versus length. The solid line is a local linear smoother, and the dashed lines are 95% confidence intervals for the average residual.

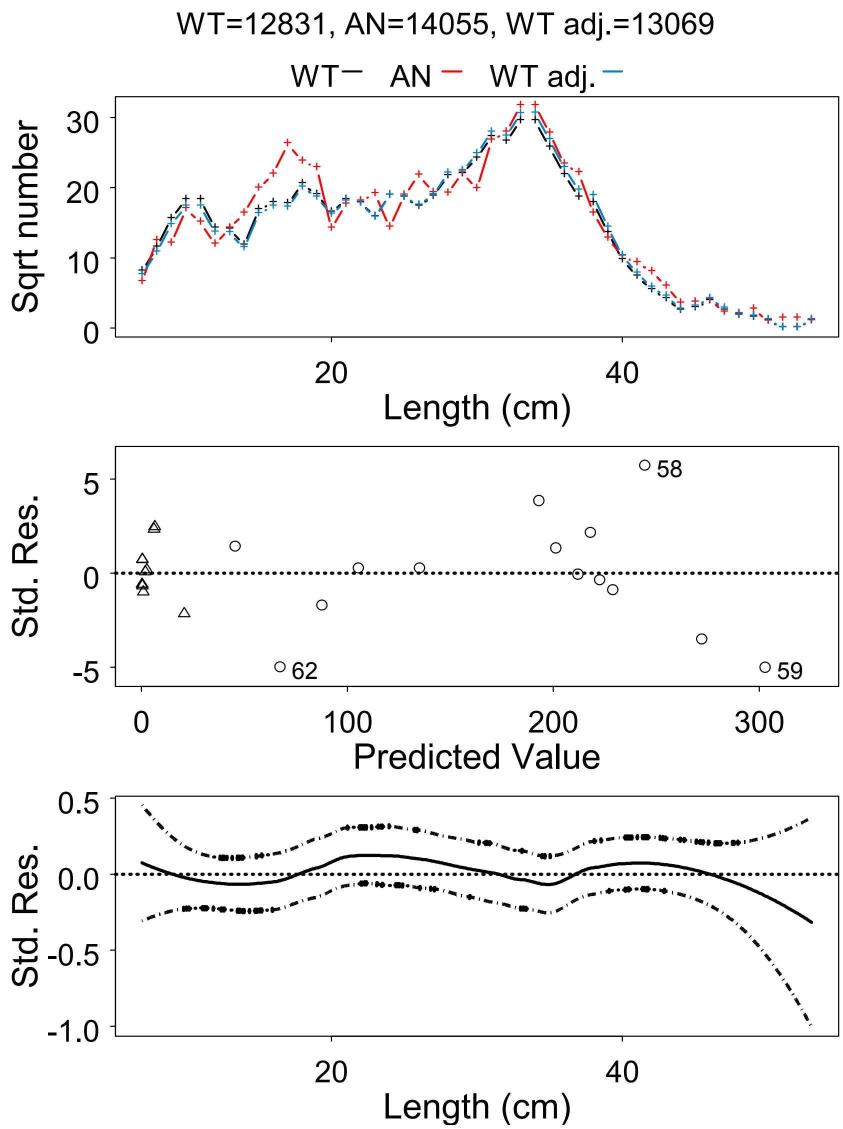

Figure 24. FE2 model results for yellowtail flounder. Total scaled catches (r^*) and WT catch adjusted by relative efficiency (ρ_l) are shown at the top. Top: Total length frequencies for all sets. Middle: Standardized (by st. err.) total chi-square residuals for each set, vs predicted WT catch. Plotting symbols: $\circ$ - Fall survey; $\triangle$ - spring survey. Residuals larger than ± 4 are identified by set number. Bottom: Standard chi-square residuals versus length. The solid line is a local linear smoother, and the dashed lines are 95% confidence intervals for the average residual.

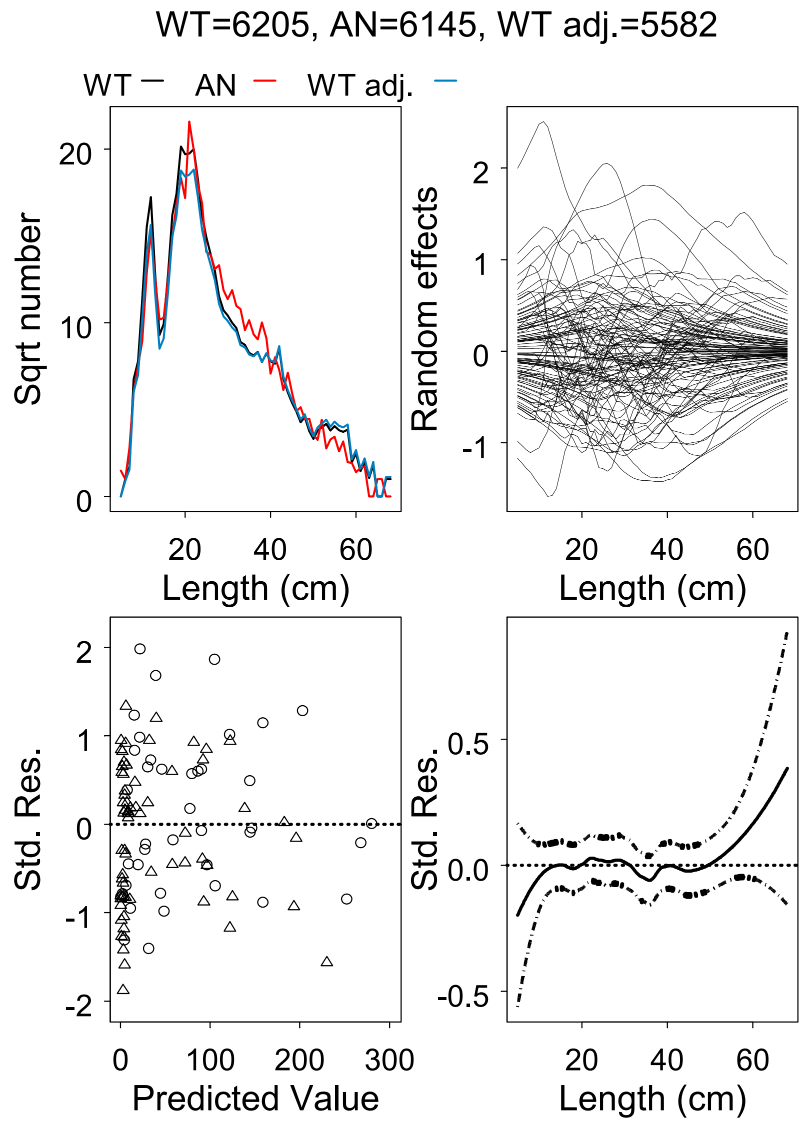

Figure 25. ME2 model results for American plaice. Total scaled catches (r^*) and WT catch adjusted by relative efficiency (ρ_l) are shown at the top. Top left: Total length frequencies for all sets. Top right: Predicted random effects, $\tilde{\delta}_{il}$, vs length, for each set. Bottom left: Standardized (by st. err.) total conditional chi-square residuals for each set, vs predicted WT catch. Plotting symbols: ○ - Fall survey; △ - spring survey. Bottom right: Standard conditional chi-square residuals versus length. The solid line is a local linear smoother, and the dashed lines are 95% confidence intervals.

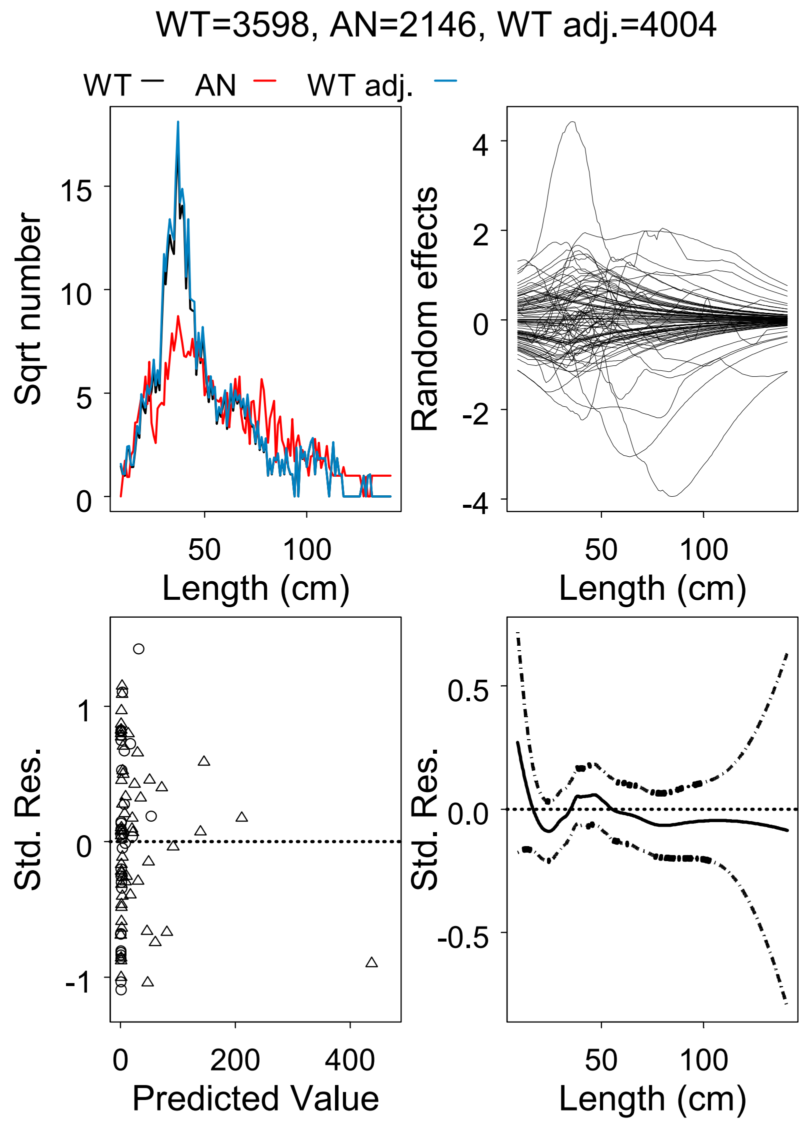

Figure 26. ME2 model results for Atlantic cod. Total scaled catches (r^*) and WT catch adjusted by relative efficiency (ρ_l) are shown at the top. Top left: Total length frequencies for all sets. Top right: Predicted random effects, $\tilde{\delta}_{il}$, vs length, for each set. Bottom left: Standardized (by st. err.) total conditional chi-square residuals for each set, vs predicted WT catch. Plotting symbols: ○ - Fall survey; △ - spring survey. Bottom right: Standard conditional chi-square residuals versus length. The solid line is a local linear smoother, and the dashed lines are 95% confidence intervals.

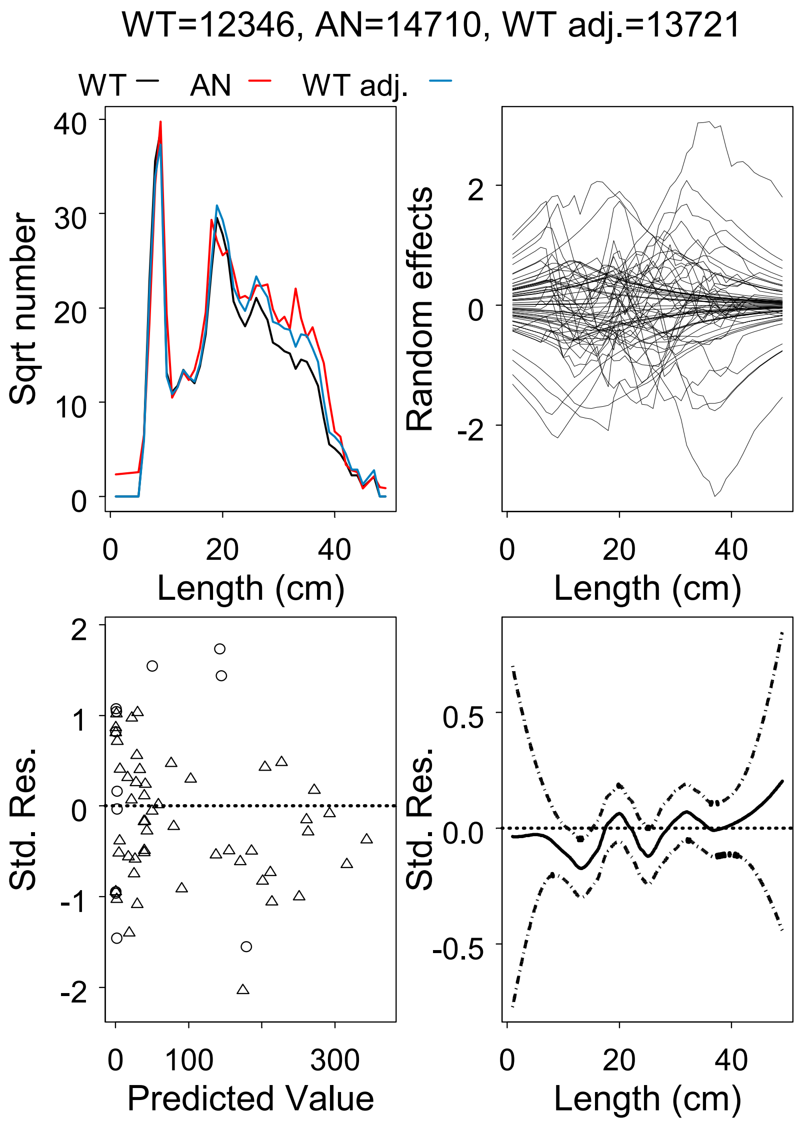

Figure 27. ME2 model results for deepwater redfish. Total scaled catches (r^*) and WT catch adjusted by relative efficiency (ρ_l) are shown at the top. Top left: Total length frequencies for all sets. Top right: Predicted random effects, $\tilde{\delta}_{il}$, vs length, for each set. Bottom left: Standardized (by st. err.) total conditional chi-square residuals for each set, vs predicted WT catch. Plotting symbols: ○ - Fall survey; △ - spring survey. Bottom right: Standard conditional chi-square residuals versus length. The solid line is a local linear smoother, and the dashed lines are 95% confidence intervals.

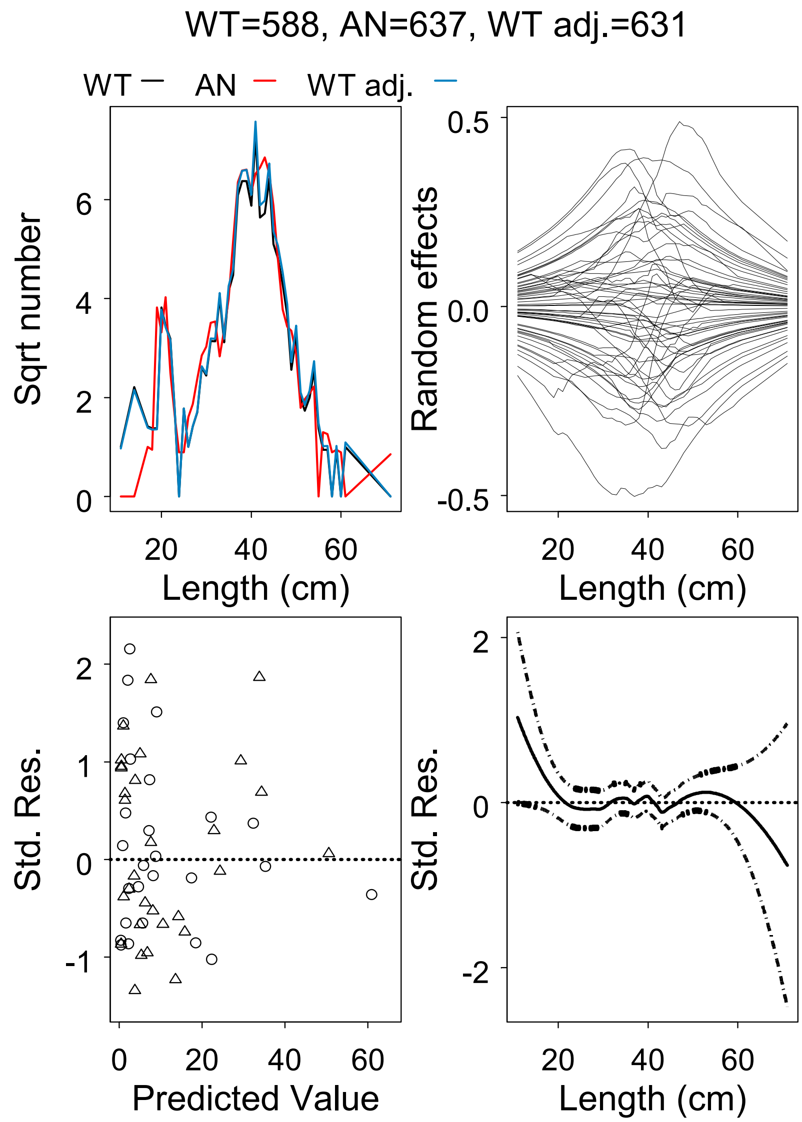

Figure 28. ME2 model results for Greenland halibut. Total scaled catches (r^*) and WT catch adjusted by relative efficiency (ρ_l) are shown at the top. Top left: Total length frequencies for all sets. Top right: Predicted random effects, $\tilde{\delta}_{il}$, vs length, for each set. Bottom left: Standardized (by st. err.) total conditional chi-square residuals for each set, vs predicted WT catch. Plotting symbols: ○ - Fall survey; △ - spring survey. Bottom right: Standard conditional chi-square residuals versus length. The solid line is a local linear smoother, and the dashed lines are 95% confidence intervals.

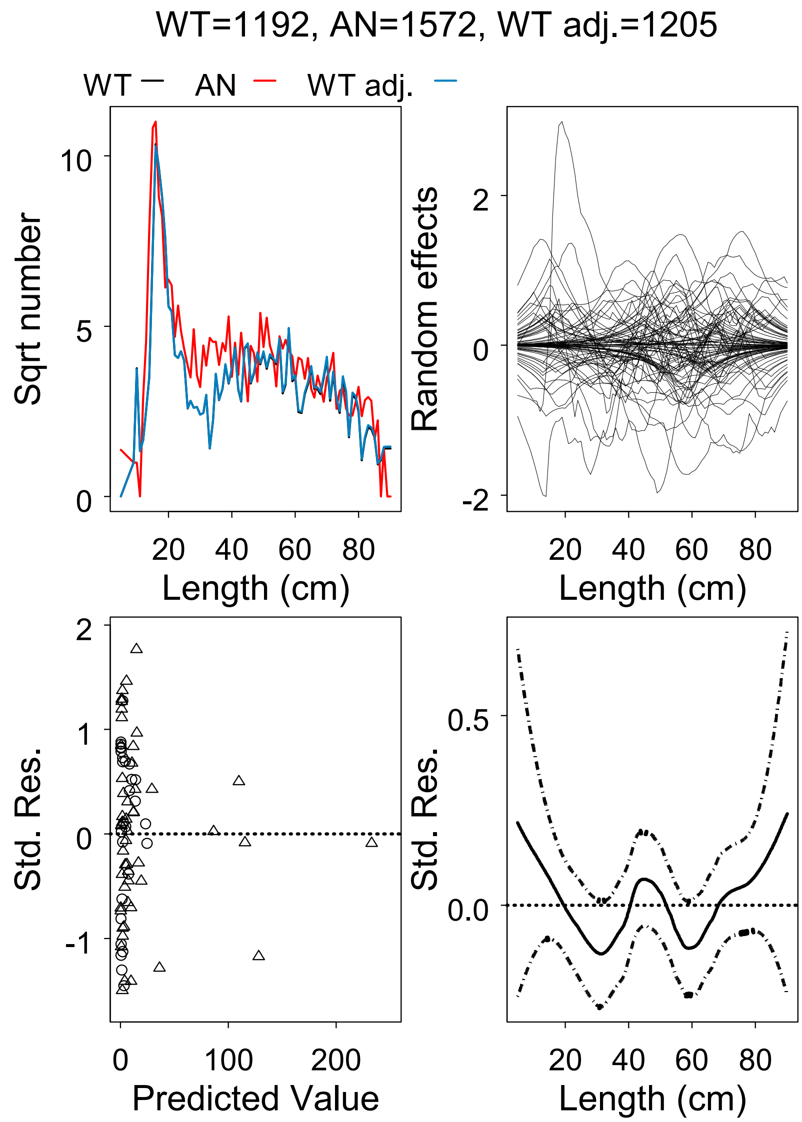

Figure 29. ME2 model results for thorny skate. Total scaled catches (r^*) and WT catch adjusted by relative efficiency (ρ_l) are shown at the top. Top left: Total length frequencies for all sets. Top right: Predicted random effects, $\tilde{\delta}_{il}$, vs length, for each set. Bottom left: Standardized (by st. err.) total conditional chi-square residuals for each set, vs predicted WT catch. Plotting symbols: ○ - Fall survey; △ - spring survey. Bottom right: Standard conditional chi-square residuals versus length. The solid line is a local linear smoother, and the dashed lines are 95% confidence intervals.

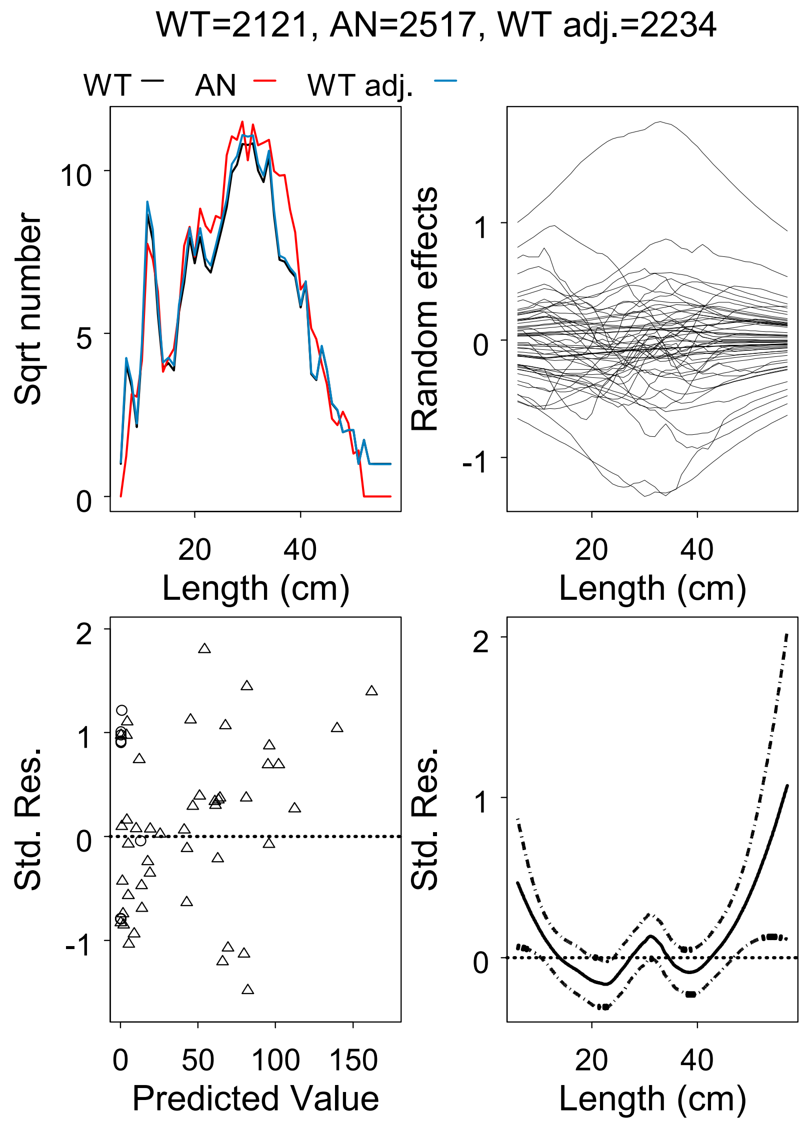

Figure 30. ME2 model results for witch flounder. Total scaled catches (r^*) and WT catch adjusted by relative efficiency (ρ_l) are shown at the top. Top left: Total length frequencies for all sets. Top right: Predicted random effects, $\tilde{\delta}_{il}$, vs length, for each set. Bottom left: Standardized (by st. err.) total conditional chi-square residuals for each set, vs predicted WT catch. Plotting symbols: ○ - Fall survey; △ - spring survey. Bottom right: Standard conditional chi-square residuals versus length. The solid line is a local linear smoother, and the dashed lines are 95% confidence intervals.

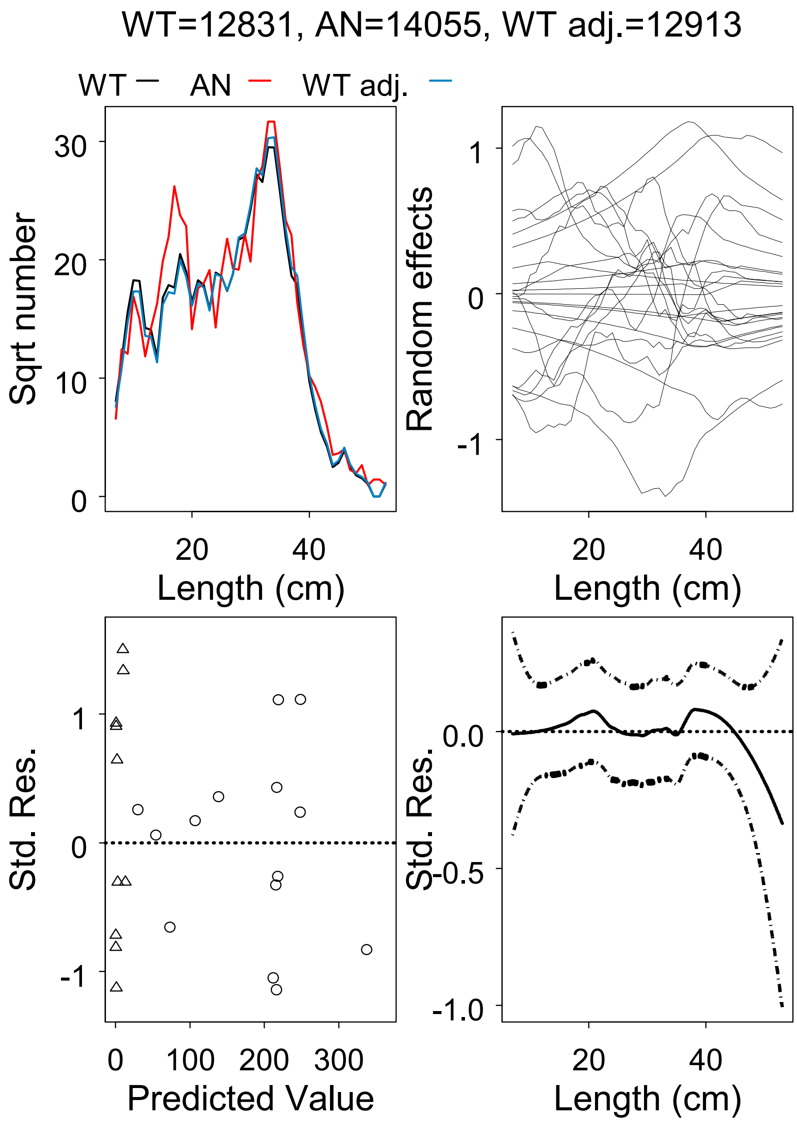

Figure 31. ME2 model results for yellowtail flounder. Total scaled catches (r^*) and WT catch adjusted by relative efficiency (ρ_l) are shown at the top. Top left: Total length frequencies for all sets. Top right: Predicted random effects, $\tilde{\delta}_{il}$, vs length, for each set. Bottom left: Standardized (by st. err.) total conditional chi-square residuals for each set, vs predicted WT catch. Plotting symbols: $\circ$ - Fall survey; $\triangle$ - spring survey. Bottom right: Standard conditional chi-square residuals versus length. The solid line is a local linear smoother, and the dashed lines are 95% confidence intervals.

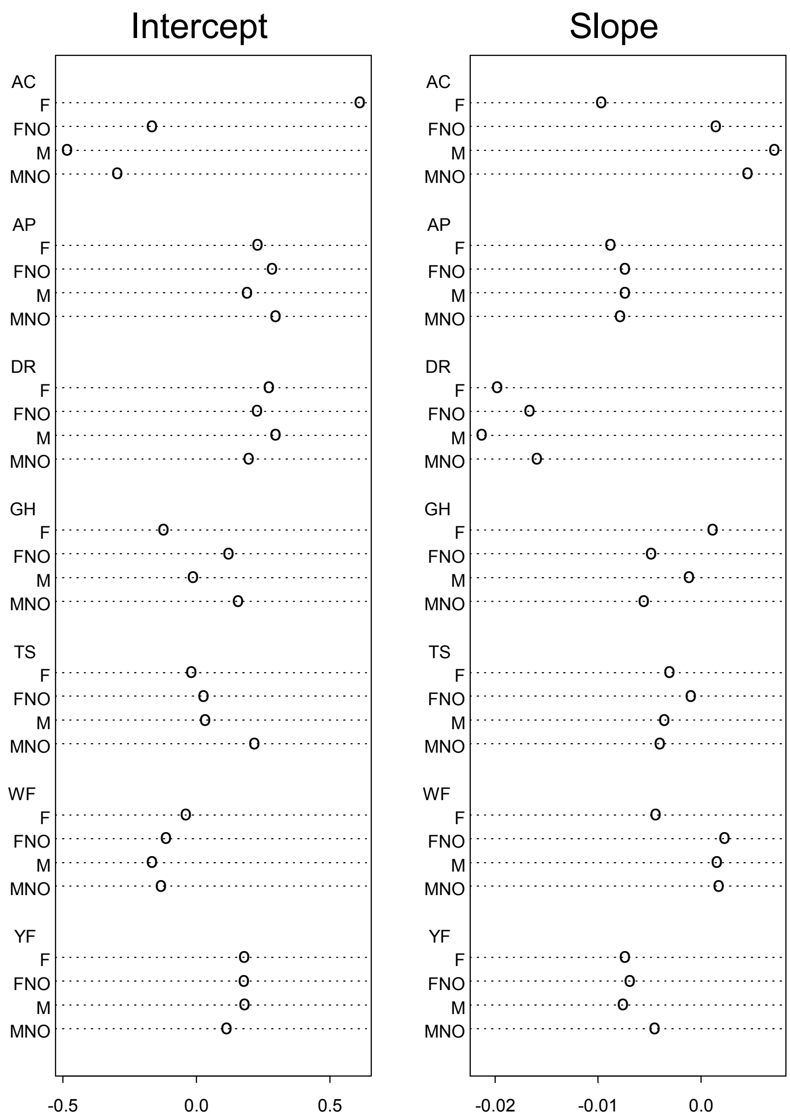

Figure 32. Estimates of $\hat{\beta}_o$ and $\hat{\beta}_1$ from the FE2 (F) and ME2 (M) models, and models with 2 potential outliers removed (FNO, MNO). Species codes: AP - American plaice; AC - Atlantic cod; DR - deepwater redfish; GH - Greenland halibut; TK - thorny skate; WF - witch flounder; YF - yellowtail flounder.

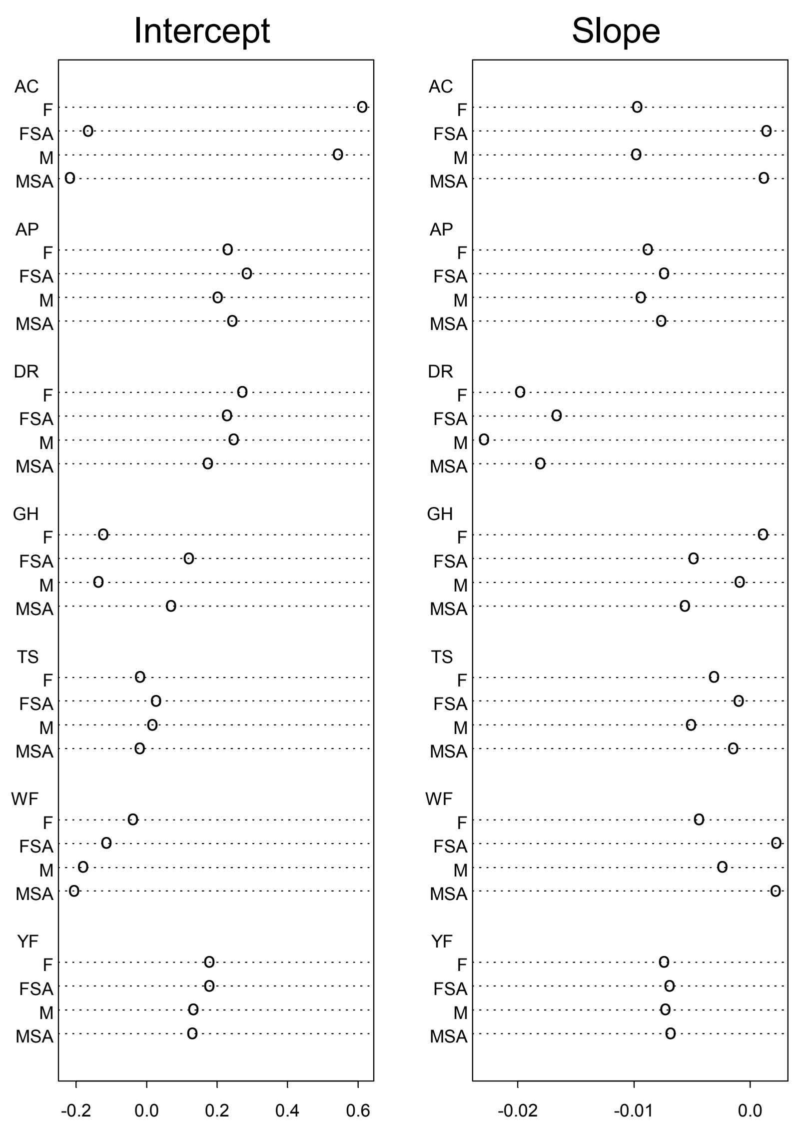

Figure 33. Estimates of $\hat{\beta}_o$ and $\hat{\beta}_1$ from the FE2 (F) and ME2 (M) models, and models with swept area standardizations (FSA, MSA). Species codes: AP - American plaice; AC - Atlantic cod; DR - deepwater redfish; GH - Greenland halibut; TK - thorny skate; WF - witch flounder; YF - yellowtail flounder.

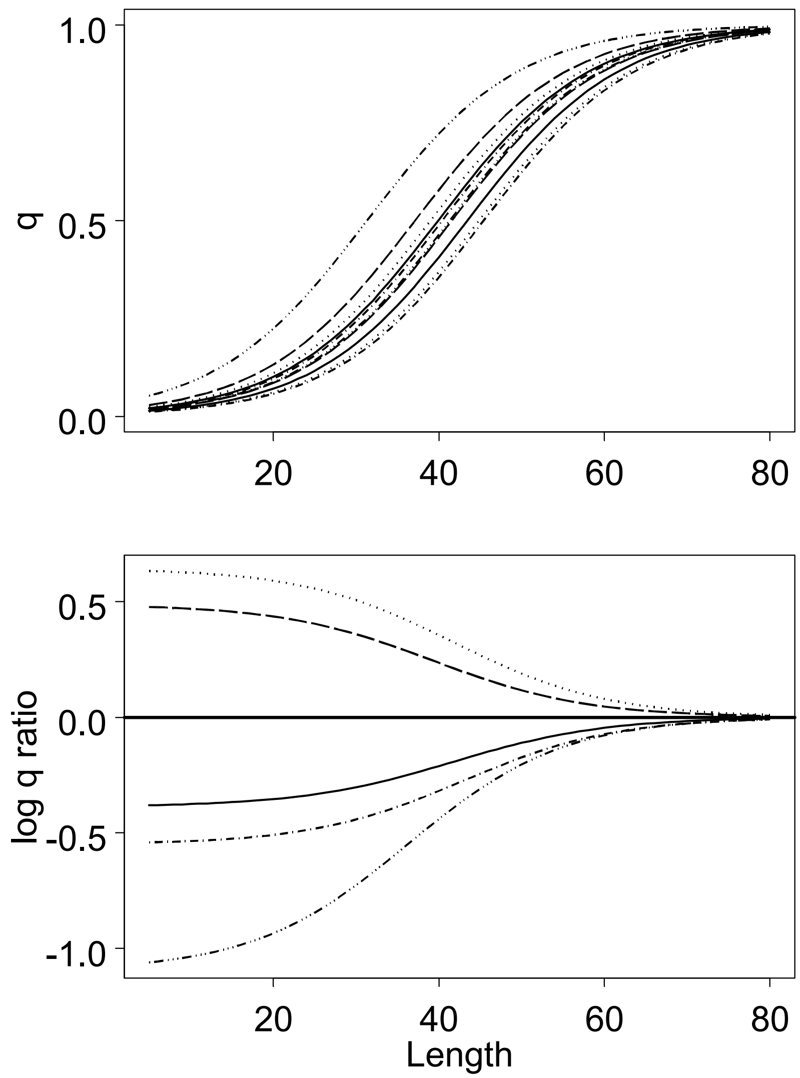

Figure 34. Top panel: Randomly generated examples of catchability (q_l) curves versus length. Two curves were generated for each of 5 tow pairs, and plotted with the same line type. Bottom panel: The log ratio of q_l for each pair, i.e. $\log(\rho_l)$. The line types correspond to those in the top panel. The solid line at zero is for reference.