A hierarchical model for relative catch efficiency from gear selectivity and calibration studies

Timothy J. Miller
Northeast Fisheries Science Center
National Marine Fisheries Service
166 Water Street, Woods Hole, MA 02543 USA
tel: +1 508 495 2365 fax: +1 508 495 2393 e-mail: timothy.j.miller@noaa.gov
Abstract

Studies of gear selectivity and calibration provide important inferences on relative catch efficiency of two or more gears. These studies can result in reduction of both commercial discards and protected species bycatch and improved stock assessments that use data obtained from multiple vessels and gears. I propose the use of a Dirichlet-multinomial model for data from either selectivity or calibration experiments where replicates consist of gears fished together as $K$-tuplets of hauls. The Dirichlet-multinomial is a hierarchical model that accounts for heterogeneity in the proportions captured by each gear. I also investigate the use of orthogonal polynomials and thin-plate regression splines as smoothers to relate relative catch efficiency and dispersion to size of fish. I demonstrate the approach with data from a study to calibrate new and retired bottom trawl survey vessels for the Northwest Atlantic. I fit a suite of models assuming different smoothers to data and provide results for several species with more detailed results for Acadian redfish. Models employing orthogonal polynomials performed better with regard to $\text{AIC}_c$ for all species, but models with regression splines performed nearly as well for some species. This is an appealing alternative to more common approaches using generalized linear mixed models because the marginal model has closed form and can easily be programmed and optimized on its own or in combination with likelihoods for other data sources in an integrated stock assessment model.
Introduction

Fishery-independent surveys are an important source of information for stock assessments worldwide. These surveys provide information on relative abundance and, when length and age frequency data are collected, productivity and mortality parameters can be inferred. Many surveys are conducted annually using the same vessel and gear from year to year whereas others may use multiple vessels and gears simultaneously. In the latter case, it is immediately apparent that differences between vessels will likely result in differences in catch efficiency that need to be considered when modeling the relationship of these data to the population in an assessment model. However, even when the annual survey is conducted using the same vessel, deterioration of equipment and technological advances result in periodic changes in vessels and gear (e.g., Byrne and Fogarty 1985; Byrne and Forrester 1991). These periodic changes lead to time series of relative abundance information with abrupt disparities in how the data relate to the population. Therefore, experiments to infer both the relative efficiency and selectivity of different vessel-gear combinations are important for almost any fishery-independent survey (Pope et al. 1975; Anonymous 1992).

Experiments and models to estimate relative catch efficiency (also referred to as fishing power correction, calibration or conversion factors) of different vessels or gears are varied (Pelletier 1998; Wilderbuer et al. 1998; Lewy et al. 2004) as are those for gear-selectivity (Pope et al. 1975; Millar and Fryer 1999). Some authors have noted their commonalities (Quinn and Deriso 1999; Fryer et al. 2003), but analytical approaches for the two problems appear to have developed independently (cf., Pelletier 1998; Millar and Fryer 1999).

I propose a Dirichlet-multinomial probability model for analyzing data from both types of experiments. The proposed model is hierarchical in nature and in the spirit of other random effects models for relative catch efficiency or gear selectivity (e.g., Fryer et al. 2003; Cadigan and Dowden 2010). However, the probability distribution can model data from any number of gears fished together as a $K$-tuplet ($K = 2$ for paired hauls) and has closed form which lends itself to incorporation of these data into likelihood-based assessment models.
Hierarchical model for relative catch efficiency

with other data sources. Within this modeling framework, I consider two general classes of
smoothers (orthogonal polynomials and penalized thin-plate regression splines) to relate size
of fish to relative catch efficiency and to the dispersion parameter that measures variability
in relative catch efficiency. I also demonstrate that the functional forms used for relative
catch efficiency in various types of gear selectivity and calibration experiments are special
cases. I apply the proposed model to estimate (size-specific) relative catch efficiency of a new
research vessel, the NOAA Ship Henry B. Bigelow, to a retired vessel, the R/V Albatross
IV, for several species that are economically important in the northeastern North America.
I describe diagnostics and results in more detail for Acadian redfish (Sebastes fasciatus).

Methods

General model for gear comparison

Let gear be defined as both the vessel and fishing gear collectively and define all $K$ hauls
using various gears conducted together in time and space, a $K$-tuplet, as occurring at station
$i$. Many of the models for gear-comparison and size-selectivity experiments have the simple
assumption that the catches measured in numbers by fishing operation $k$ at station $i$, $C_{ik}(L)$,
have expected value

$$E [C_{ik}(L) | X_{ik}(L)] = q_{ik}(L)X_{ik}(L)$$

and covariance of gears $k$ and $l$,

$$Cov [C_{ik}(L), C_{il}(L) | X_{ik}(L), X_{il}(L)] = \sigma_{ikl}(L).$$

As parameterized here, the expected number caught is a simple relationship between what
I will call catch efficiency $q_{ik}(L)$, and the subpopulation of the stock (in number of fish) in
proximity to the gear at station $i$, $X_{ik}(L)$. I make explicit here the assumption that the
mean and variance (and covariance) can change with size as measured by length $(L)$, but
other covariates hypothesized to be important predictors of catch, such as depth and season (e.g., Benoît and Swain 2003), could be similarly incorporated. The general model (eqs. 1-2) is identical to the model proposed by Pelletier (1998, eq. 2) for gear calibration (without specifying a relationship to length).

The model described by (Millar and Fryer 1999, eq. 5) for gear selectivity can also be shown to be a special case of eqs. 1-2. The models for gear selectivity studies commonly focus specifically on “contact” selectivity of gear \( k \), \( r_k(L) \), which is the proportion of fish captured of those that contact the gear (Millar 1992; Millar and Fryer 1999). Let \( a_{ik}(L) \) and \( c_{ik}(L) \) be the probability of a fish in the subpopulation being available to gear \( k \) and the probability of contacting the gear given it is available at station \( i \), respectively. Therefore, \( q_{ik}(L) = a_{ik}(L)c_{ik}(L)r_{ik}(L) \). Also, by definition in Millar and Fryer (1999), \( p_{ik}(L)\lambda_i(L) = \lambda_{ik}(L) \) is the expected number of fish that contact gear \( k \) at station \( i \). Then \( \lambda_{ik}(L) = a_{ik}(L)c_{ik}(L)X_{ik}(L) \) and the expected number caught from eq. 5 in Millar and Fryer (1999) and eq. 1 here are equivalent because

\[
r_{ik}(L)p_{ik}(L)\lambda_i(L) = \frac{q_{ik}(L)}{a_{ik}(L)c_{ik}(L)} a_{ik}(L)c_{ik}(L)X_{ik}(L) = q_{ik}(L)X_{ik}(L).
\]

The model in eq. 5 of Millar and Fryer (1999) specifies all parameters to be constant across stations, but the analogous model that allows variation in the parameters (e.g., Fryer et al. 2003) is also a special case of eqs. 1-2 because both require

\[
V[C_{ik}(L)|X_{ik}(L)] = \sigma_{ikk} = E[C_{ik}(L)|X_{ik}(L)]
\]

due to the Poisson assumption, and all catches to be independent. In fact, I specify that parameters may be both station- and gear-specific in eqs. 1-2 so that variability among stations can be considered. Heterogeneity in the catch efficiency (at size) across hauls has long been known to occur (Fryer 1991) and can arise in many ways. Variation in the towing operation, unconsidered environmental factors (e.g., time of day, substrate type, time of
year), and spatial and temporal variation in behavior of the species all can cause the catch efficiencies and, therefore, the ratios of catch efficiencies of different gears to vary across stations.

A Dirichlet-multinomial model

Similar to previous studies, I assume that the number of fish (at size) caught at station \( i \) by gear \( k \) are Poisson random variables (Millar 1992; Lewy et al. 2004; Cadigan and Dowden 2010). The catches are assumed independent conditional on densities of fish in the tow path of each gear \( D_{ik}(L) \) with mean \( E[C_{ik}(L)|D_{ik}(L)] = q_{ik}(L)f_{ik}(L)A_{ik}D_{ik}(L) \). The fraction of the catch sampled and the area or volume swept are \( f_{ik}(L) \) and \( A_{ik} \), respectively. I parameterize \( X_{ik}(L) \) as the product of a swept area (or volume) and density of fish here because of the focus on trawl gear (Paloheimo and Dickie 1964) and the sampling fraction may be size-dependent, but I assume this and the swept area are known.

A well known attribute that many have exploited is that conditional on the catch by all gears \( C_i(L) \), the numbers caught by each gear follow a multinomial distribution with probabilities of capture

\[
p_{ik}(L) = \frac{E[C_{ik}(L)|D_{ik}(L)]}{\sum_{l=1}^{K} E[C_{il}(L)|D_{il}(L)]}
\]

(Millar 1992; Lewy et al. 2004). In gear selectivity and calibration studies, the variability in relative catch efficiency among stations has usually been modeled using generalized linear mixed effects models where parameters that vary across stations are assumed normally distributed. The methods for fitting these models have improved significantly (cf. Fryer 1991; Holst and Revill 2009; Cadigan and Dowden 2010). However, fitting algorithms are complicated due to the need to either numerically integrate or approximate the integral of the likelihood function over the random effects (e.g., Diggle et al. 2002, pp. 172-175).

I propose an alternative mixed effects model. Assume the heterogeneity in probabilities \( \mathbf{p}_i(L) = (p_{ik}(L), \ldots, p_{ik}(L))' \) can be described by a Dirichlet distribution which is a multivariate distribution of dimension \( K \) with mean \( \boldsymbol{\pi}_i(L) = (\pi_{i1}(L), \ldots, \pi_{iK}(L))' \) and covariance
Hierarchical model for relative catch efficiency

\[
V[p_i(L)] = \frac{1}{\phi(L) + 1} \left[ \text{diag} \{ \pi_i(L) \} - \pi_i(L)\pi_i(L)' \right]
\]

where \( \phi > 0 \) is a dispersion parameter that may also be a function of covariates including length, swept area, and sampling fraction. The Dirichlet distribution is a natural choice because the probabilities of capture must take on values between 0 and 1 and negative correlation of the probabilities of capture for different gears is appropriate because they must sum to unity. I assume the mean density of fish is the same for each of the gears at a given station and mean catch efficiency for gear \( k \) is the same across stations so that mean capture probabilities are not dependent on the densities at each station,

\[
\pi_{ik}(L) = \frac{q_k(L)f_{ik}(L)A_{ik}D_i(L)}{\sum_{l=1}^{K} q_l(L)f_{il}(L)A_{il}D_i(L)} = \frac{q_k(L)f_{ik}(L)A_{ik}}{\sum_{l=1}^{K} q_l(L)f_{il}(L)A_{il}}.
\]

In gear calibration and some gear selectivity studies where gears are not fished simultaneously, it is common that catch efficiencies of all gears are unknown so that we can only estimate catch efficiency of the \( K - 1 \) gears relative to a chosen reference gear. In such case, an estimable parameterization for the mean probability of capture for gears 1, \ldots, \( K - 1 \) is

\[
\pi_{ik}(L) = \frac{\rho_k(L)f_{ik}(L)A_{ik}}{f_{iK}(L)A_{iK} + \sum_{l=1}^{K-1} \rho_l(L)f_{il}(L)A_{il}}
\]

where the reference gear is \( K \) and \( \rho_k(L) = q_k(L)/q_K(L) \) is the relative catch efficiency (at length) of gear \( k \).

When the probabilities of capture arise from a Dirichlet distribution, the vector of catches by each gear \( C_i(L) = (C_{i1}(L), \ldots, C_{iK}(L))' \) conditional on the total catch at the station \( C_i(L) \) has a Dirichlet-multinomial (D-m) distribution with mean \( E(C_i(L)|C_i(L)) = \)
Hierarchical model for relative catch efficiency

\[ C_i(L) \pi_i(L), \text{ and covariance matrix} \]

\[(6) \quad V(C_i(L)|C_i(L)) = C_i(L) \left[ \text{diag} \{ \pi_i(L) \} - \pi_i(L)\pi_i(L)' \right] \frac{\phi_i(L) + C_i(L)}{\phi_i(L) + 1}. \]

Note that for large \( \phi \), the covariance matrix of the Dirichlet distribution (eq. 3) goes to zero and that of the D-m (eq. 6) approaches that of the simpler multinomial. Also, for the special case of \( K = 2 \) gear types, the Dirichlet is equivalent to the more commonly known beta distribution and the marginal distribution of the catch by one gear conditional on the total at station \( i \) is beta-binomial.

An alternative motivation for the Dirichlet distribution that could be considered is based on assuming the mean number captured by each gear \( E[C_{ik}(L)|D_{ik}(L)] \) are independent gamma random variables with the same scale parameter (Stuart and Ord 1994, pg. 271).

Let

\[ E[C_{ik}(L)|D_{ik}(L)] \sim \text{Gamma} \left( \tau_i(L)q_k(L)f_{ik}(L)A_{ik}D_i(L), \frac{1}{\tau_i(L)} \right) \]

where the mean is \( q_k(L)f_{ik}(L)A_{ik}D_i(L) \) and \( 1/\tau_i(L) \) is the scale parameter. The mean vector for the Dirichlet is still that given in eq. 4, but the dispersion parameter \( \phi \) now is also implicitly a function of the means of the gamma random variables and the scale parameter,

\[ \phi_{ik}(L) = \tau_i(L)D_i(L)q_K(L) \left[ f_{iK}(L)A_{iK} + \sum_{t=1}^{K-1} \rho_t(L)f_{it}(L)A_{it} \right]. \]

This is an over-parameterized and inestimable D-m model because the unknown catch efficiency of the reference gear, density, and scale parameters are confounded, and their product is station-specific. If the gamma random variables are assumed to describe the variability in catches across all stations, then the scale and density parameters are constant across stations, so that \( E[C_{ik}(L)] = g_k(L)f_{ik}(L)D(L) \) and

\[(7) \quad \phi_{ik}(L) = \alpha(L) \left[ f_{iK}(L)A_{iK} + \sum_{t=1}^{K-1} \rho_t(L)f_{it}(L)A_{it} \right]. \]
where $\alpha(L) = \tau(L)D(L)q_K(L)$. This parameterization (eq. 7) results in a D-m model with the same probability vector (eq. 5) and although the unknown scale, density, and catch efficiency parameters are confounded, their product $\alpha(L)$ is estimable. Interestingly, the catches $C_{ik}$ are also implicitly independent (conditional on $D_i$ and not on $C_i$) and marginally have a negative binomial distribution.

Due to the relationship of the D-m and multinomial models, a logit transformation of the vector of proportions is a natural way to model the relationship to size of the fish. Using the multivariate logit link with gear $K$ as the base-line category (Agresti 2002, pp. 267-268),

$$\log \left( \frac{\pi_{ik}(L)}{\pi_{iK}(L)} \right) = \log [\rho_k(L)] + \log \left[ \frac{f_{ik}(L)}{f_{iK}(L)} \right] + \log \left( \frac{A_{ik}}{A_{iK}} \right),$$

which is a linear function of the logs of the ratios of sampling fractions and swept areas as well as the log of the relative catch efficiency. In the case of $K = 2$, eq. 8 is the same as the usual logit transformation used in binomial models for gear comparison and calibration (e.g., Fryer et al. 2003; Holst and Revill 2009).

I use eq. 8 in the example application for relative catchability and discuss models for $\rho(L)$ in the following section.

**Models for size-based relative catch efficiency**

In stock assessment models, the catch efficiency-at-size $q_{ik}(L)$ is usually parameterized as the product of a fully-selected catch efficiency $q_k$ and the selectivity function $s_k(L)$ (e.g., Quinn and Deriso 1999). The expected number caught conditional on the number available is then $E(N_{ki}(L)) = q_k s_k(L) A_{ki} D_{ki}(L)$. When trawl gear is being investigated and contact selectivity is of interest, one might assume both operations have logistic selectivity (Millar and Fryer 1999). However, the logistic selectivity assumption for trawl gear may not hold for many species when the numbers caught are also a function of availability (Huse et al. 1999; Jacobson et al. 2001; Butterworth and Rademeyer 2008). Availability is an issue in calibration studies and some experimental designs for gear selectivity where paired hauls are
Hierarchical model for relative catch efficiency

not conducted simultaneously in space and time (Fryer et al. 2003).

In some gear selectivity studies using trouser trawls, one “leg” of the trouser trawl (reference gear $K$) is usually assumed to retain all fish that contact the gear and the probabilities of contacting each gear given availability are assumed to be complimentary $c_k(L) = 1 - c_K(L)$ (e.g., Cadigan and Millar 1992; Millar and Walsh 1992). The contact selectivity of the others can theoretically be estimated under these assumptions. It might be safe to assume that the probability of being availability and contacting each gear is proportional $(a_k(L)c_k(L) = \beta a_K(L)c_K(L))$ given the construction of the trouser trawl if the proportions available and contacting each gear are simply a function of the sizes of the openings of each leg of the trouser trawl. In such case,

$$\rho_k(L) = \frac{a_k(L)c_k(L)r_k(L)}{a_K(L)c_K(L)r_K(L)} = \beta r_k(L)$$

where $\beta/(1 + \beta)$ would be equivalent to the proportion contacting gear $k$, also known as a “split” parameter (cf. Millar and Fryer 1999). Similarly, when covered codends are used (e.g., Reeves et al. 1992; Zuur et al. 2001), the cover (gear $K$) is assumed to retain all fish that pass through the mesh (gear $k$) and $a_k(L)c_k(L) = a_K(L)c_K(L)$ because gears $k$ and $K$ are fished on the same haul which results in

$$\rho_k(L) = \frac{a_k(L)c_k(L)r_k(L)}{a_K(L)c_K(L)r_K(L)} = \frac{r_k(L)}{1 - r_k(L)}.$$

In cases where availability must also be considered, we may not wish to assume that the selectivity has a specific functional form and an alternative is to estimate a smoother that allows a wide range of flexibility in the relationship of size and relative catch efficiency. A smoother is particularly useful when calibrating different gears where prediction of the relative catch efficiency rather than testing gear effects is of primary importance.

There are a wide variety of smoothers that could be used, but, in general, a smoother for
(log) relative catch efficiency is linear in functions of length,

\[
\log [\rho_{ik}(L)] = \sum_{r=0}^{R} \beta_r g_r(L).
\]

In eq. 9, \( R \) is the number of terms, \( g_r(L) \) is the \( r \)th function of length and \( \beta_0, \ldots, \beta_R \) are the coefficients to be estimated. For example, Lewy et al. (2004) fitted a polynomial model for relative catch efficiency where \( g_r(L) = L^r \) and \( r \) is the degree of the polynomial. Holst and Revill (2009) used orthogonal polynomials where \( g_r(L) \) is a polynomial of degree \( r \) in \( L \) that is orthogonal to all \( g_{r'}(L) \) for \( r \neq r' \) (Kennedy and Gentile 1980, pp. 342-347). Orthogonal polynomials are generally better than ordinary polynomials in estimation procedures due to the lack of correlation of the terms (Wood 2006, pg. 305). Smoothers such as those used in fitting generalized additive models (GAMs) are another option (Hastie and Tibshirani 1990). Fryer et al. (2003) used a loess smoother to model the relationship of selectivity and relative catch efficiency to length, but they had to make assumptions on the degree of smoothing.

There has been substantial advancement in methods for fitting GAMs including the ability to estimate the degree of smoothing for certain penalized regression spline smoothers (e.g., Wood 2000). Furthermore, the work by Rigby and Stasinopoulos (2005) and Stasinopoulos and Rigby (2007) has provided a means of fitting smoothers of covariates to scale and shape as well as the mean of various distributions outside of the exponential family. In the present context, this allows us to consider smoothers of length for the dispersion parameter of the D-m model.

**Calibration of the Henry B. Bigelow and Albatross IV**

In 2008, an extensive paired tow experiment was conducted to provide a means to estimate the catch efficiency of a new research vessel, the Henry B. Bigelow, relative to the old vessel, the Albatross IV, for various species managed by the Northeast Fisheries Science Center (NEFSC). The Albatross IV was used for nearly half of a century (1963-2008) by the NEFSC to conduct annual bottom-trawl surveys during the spring and fall in the Northwest
Hierarchical model for relative catch efficiency

Atlantic Ocean (Azarovitz 1981). These data are integral to management of many exploited
groundfish, pelagic, and invertebrate stocks. The design of the calibration experiment and
the characteristics of the Albatross IV and Henry B. Bigelow are described in detail elsewhere
(NEFSC Vessel Calibration Working Group 2007), but for the purposes of this work note
that the Henry B. Bigelow is larger, faster, and quieter than the retired Albatross IV and
the trawl gear on the Henry B. Bigelow has different doors and a net with greater wingspan,
headrope height, and different meshsizes in the panels and codend. Furthermore, there are
differences in method and duration of trawl deployment and in the processing of catches
after the tow is retrieved. Taken together, all of these differences suggest an unpredictable
effect of fish size to the relative catch efficiency.

Paired tows were made by the Albatross IV and Henry B. Bigelow at 636 stations during
both the traditional spring and fall NEFSC survey seasons as well as the summer months,
June and July, to target groundfish stocks poorly sampled during the seasonal surveys. The
paired tows were designed to have a temporal and spatial offset of the individual tows at
each station so that, at least for predominant groundfish, effects of the towing by one on
the catches of the other were minimized while available densities were kept similar. Over
300 species or species groups were observed during the calibration study (Miller et al. 2010),
but for this paper I performed analyses for Acadian redfish, Atlantic cod (Gadus morhua),
black sea bass (Centropristis striata), haddock (Melanogrammus aeglefinus), summer flounder
(Paralichthys dentatus), and winter flounder (Psuedopleuronectes americanus). For illustra-
tive purposes, I give detailed results for Acadian redfish.

In these analyses, I treated the Albatross IV as the reference K. Because there are only
two gears, the D-m model reduces to the simpler beta-binomial and there is a single \( \rho_k(L) \) to
estimate. I fit models where penalized thin-plate regression splines (SP) (Wood 2003, 2006)
or orthogonal polynomials (Kennedy and Gentile 1980; Chambers and Hastie 1992) (OP)
were assumed for both the relative catch efficiency (eq. 9) and the dispersion parameter. For
OP and SP models, I assumed the dispersion could be a function of swept area and sampling
Hierarchical model for relative catch efficiency

fraction as well as the smoother of length,

\[ \log [\phi_{ik}(L)] = \alpha_0 \log \left[ \frac{A_{ik}}{A_{iK}} \right] + \alpha_1 \log \left[ \frac{f_{ik}}{f_{iK}} \right] + \sum_{r=0}^{R} \beta_r g_r(L). \]

Fitted models included those where terms for swept area (SA) and(or) sampling fraction (SF) were dropped from the dispersion model. Within the SP class, I also fit models where the relative catch efficiency and dispersion parameters were not functions of length. Lastly, I also considered the beta-binomial model where the expected catches have gamma distributions (eqs. 5-7) with constant density and scale parameters (OP-G). I assumed orthogonal polynomial models for \( \alpha(L) \) in the dispersion (eq. 7) of the OP-G models. For the OP and OP-G models, the degree of the polynomial ranged between 0 (no length effect) and 12 for the relative catch efficiency and the dispersion parameter, but some models with greater complexity than required for particular species failed to converge.

I fit all models in the statistical programming environment R and for SP and OP models I used the \texttt{gamlss}, \texttt{gamlss.add} and \texttt{mgcv} packages (Wood 2001; Stasinopoulos and Rigby 2007; R Development Core Team 2010). The \texttt{gamlss} function uses maximum penalized likelihood to estimate parameters and the degree of smoothing for the SP models. For the OP and OP-G models, the degree of smoothing is fixed parameters are estimated by maximum likelihood. I maximized programmed log-likelihood functions for the OP-G models using the \texttt{nlminb} function. I compared relative performance of fitted models using the Akaike Information Criterion with correction for small sample bias correction (AIC\(_c\)) (Hurvich and Tsai 1989).

For Acadian redfish, I also qualitatively assessed the goodness-of-fit of the best model in two ways. First, I compared the predicted \( \rho \) and \( \phi \) at length from the best performing beta-binomial model with corresponding predicted values from models fitted to each 1 cm length class without sampling fraction and swept area as covariates in the dispersion component. The latter models predict a unique \( \rho \) and \( \phi \) for each length class, analogous to a “saturated” model. The predicted values from the more parsimonious model with smoothers should
trend similar to the length-class-specific predictions. Second, I inspected plots of randomized normal quantiles of residuals for the best performing model (Dunn and Smyth 1996).

Results

Acadian redfish

Swept area was measured for both the Henry B. Bigelow and Albatross IV at 94 stations where Acadian redfish were observed by either vessel. At 23 of the 94 stations a subsample of the catch for either the Henry B. Bigelow or Albatross IV was measured for length. Total length was measured for 14,653 redfish, ranging from 4 to 43 cm (Figure 1). In all, there were 1360 total observations by station and 1 cm length class. The Henry B. Bigelow generally caught more total redfish and at a larger proportion of stations than the Albatross IV, but the differences were greatest at lengths 8-15 cm. Although very few fish were caught in the 4 cm and 43 cm size classes all were caught by the Albatross IV in the former and all by the Henry B. Bigelow in the latter.

Several of the fitted beta-binomial models for Acadian redfish performed similarly well with respect to AICc (Table 1). The top six performing models had AICc values within 2 units of each other and all of the top ten models were OP models that assumed orthogonal polynomials of fish length for both ρ and φ. The best performing models assumed similar degree of polynomials for relative catch efficiency (7-9) and dispersion (1-3) and most included either sampling fraction or swept area or both as a covariate in the dispersion portion of the model. The best OP-G and SP models performed very poorly relative to the best OP models, but the latter performed much better than the former.

Inspection of the residuals of the best performing model for Acadian redfish reveals no strong pattern with either the fitted number captured by the Henry B. Bigelow nor the total number captured at each station and there does not appear to be any strong departure from normality (Figure 2). Furthermore, the predicted relative catch efficiency and dispersion parameter at length follows the trends in the length-class-specific model.
estimates well (Figure 3). Predicted relative catch efficiency also reflect trends in the raw data. In particular note that the predicted $\rho$ approaches zero at 4 cm due to all fish captured by the Albatross IV whereas predicted $\rho$ approaches infinity at 43 cm due to all fish captured by the Henry B. Bigelow. The catch efficiency of the Henry B. Bigelow is estimated to be at least 10 times greater than that of the Albatross IV for 9-10 cm fish.

It also appears that accounting for changes in the dispersion parameter with length can be important for estimating relative catch efficiency. The OP model that has the same degree of polynomial for the relative catch efficiency as the best performing model, but models the dispersion parameter as only a function of the sampling fraction and swept area (constant with respect to length) predicts higher relative catch efficiency at the 8-11 cm length classes (Figure 3).

Other species

Among other analyzed species, all were observed at greater than 100 stations by either the Henry B. Bigelow or the Albatross IV (range 111-176) and there were over 500 total observations by station and 1 cm size class (range 511 - 2374) (Table 2). The amount of information in terms of number of stations where the species was observed, total number of observations, and total number of lengths measured was least for black sea bass and greatest for haddock. Catches were subsampled for length measurements for only three of the five species, but at very few stations except for haddock. Atlantic cod had the largest range in observed lengths (4-121 cm). As with Acadian redfish, the proportion of stations where fish were caught and total number of fish caught by vessel were generally greater for the Henry B. Bigelow than the Albatross IV for all of these species (Figure 4). However, the Albatross IV caught small black sea bass and Atlantic cod at a greater proportion of stations than the Henry B. Bigelow and in greater numbers. For haddock, Atlantic cod, and the two flounders, the proportion captured by the Albatross IV was greater in the intermediate size classes than the adjacent small and large size classes.
The OP or OP-G models performed better than SP models for all species (Table 3). For black sea bass and summer and winter flounder, OP-G models performed better than OP models, whereas the reverse was true for haddock and Atlantic cod. However, some OP and SP models performed nearly as well as the best OP-G model for black sea bass and the two flounders and one OP-G model performed well for Atlantic cod.

Based on the best performing models, there were some similarities in the way relative catch efficiency and the dispersion parameter change with fish length for certain species (Figure 5). The relationship of relative catch efficiency to length is similar for the two flounder species with greater estimates of $\rho$ at the smallest and largest size classes relative to corresponding estimates at intermediate sizes, but the magnitude of $\rho$ at some of the intermediate sizes differed between the flounders. The occurrence of two pronounced modes at small and large sizes was a commonality in patterns of relative catch efficiency for black sea bass, haddock and Atlantic cod as well.

Patterns across species were different for the dispersion parameter. Haddock, Atlantic cod, and summer flounder all exhibited greater dispersion parameter estimates at small and large size classes relative to the intermediate size classes. A decrease in dispersion parameter with length was estimated for black sea bass and a multi-modal relationship was estimated for winter flounder. However, precision in the estimated dispersion was relatively poor at the small and large sizes for all species and some intermediate sizes for winter flounder.

**Discussion**

The better performance of the OP and OP-G models than the SP models was somewhat surprising. If the smoothness penalty for the regression spline is stronger than that given by AIC$_c$ there would be fewer estimated degrees of freedom for the spline smoothers in the best SP model than the degrees of freedom used by best OP or OP-G models, but this was not consistently the case. For Acadian redfish and black sea bass, the estimated degrees of freedom for the best SP model was greater than the degrees of freedom associated with the
best OP or OP-G model whereas the reverse was true for the other species. However, the
best SP models performed nearly as well as the OP or OP-G models for the flounders and
perhaps SP models would perform better for other species I did not consider. There are also
fundamental differences between the ranges of curves that can be expressed with orthogo-
nal polynomials and thin-plate regression splines. Because many of the models performed
similarly with respect to AICc, model-averaged estimates of relative catch efficiency might
provide a better inferences made from these data (Burnham and Anderson 2002).

Current generalized linear mixed models (GLMMs) for both gear calibration and selectiv-
ity experiments assume normal random effects in the link transformation of the conditional
mean model. For example, Fryer et al. (2003) assumes that the parameters defining the
relationship of selectivity or catch rate at each haul are normally distributed across paired
hauls whereas the random effects model that Cadigan and Dowden (2010) propose for rela-
tive catch efficiency assumes that the difference in the log-densities is normally distributed.
In this paper, I proposed an alternative mixed effects model for these studies that assumes
the conditional probability of capture is Dirichlet distributed. The proposed D-m model has
closed form unlike the GLMM model and multiple gears \( K > 2 \) are easily compared, both
of which can be desirable properties.

The closed form of the D-m model is particularly appealing when the data from gear
selectivity and gear comparison experiments are incorporated with other survey and fishery
data in stock assessment models, which is often a primary reason that these studies are
undertaken. Although it is more common to perform analyses of these data and use the
resulting estimates in assessment models, simultaneous estimation of the parameters of the
fish population and gear selectivity/comparison models allows more accurate estimation
of precision of the estimated population parameters. The ability to write programs that
calculate the likelihood of the D-m model is also useful apart from stock assessment when
more than two gears are compared. The multivariate generalization of the alternative GLMM
approach is a multinomial mixed effects model and software does exist to fit these models
(e.g. Hedeker 1999), but programming a D-m model is relatively simple and it allows freedom to consider a wide range of functional forms for the relationship of length or other covariates to relative catch efficiency and dispersion parameters.

I also proposed a OP-G model that assumes the mean numbers (at length) captured by each gear at a station were gamma distributed. Regression spline smoothers could theoretically be used to model the relationships of $\rho$ and $\alpha$ to length instead of the orthogonal polynomials in the OP-G models, but the algorithms available in the \texttt{gamlss} package would have to be modified to fit these models. Although these models provided the best fits for black sea bass and the two flounders, the assumption that the gamma random variables must represent variability across stations may be undesirable. Furthermore, there are other, perhaps, more realistic models that would allow the gamma random variables to instead be correlated rather than identical in value. Various probability models for multivariate gamma distributed random variables could be considered (e.g., Kibble 1941; Downton 1970; Smith et al. 1982), but the form of the marginal joint distribution of the catches would likely not be closed and more complex computational methods would be required.

A common application of relative catch efficiency estimates is in calibrating catches by one gear to be on the same scale as that of another. When the estimates are size-specific, there is a concern that sizes will eventually be observed outside the range observed during the comparison study. Although caution must be used in making predictions at these unobserved sizes, they are necessary to integrate relative abundance data from different gears. Polynomial model predictions will be biased, but regression spline smoothers may be parameterized to make these extrapolations in various ways that can be deemed justifiable by the investigator.

Because relative catch efficiencies appear to depend, for many species, on the size structure of the population being sampled, relative catch efficiency estimates that are not a function of size will be unreliable when applied to survey catches that sample different size structures. I also found that the estimated relationship of relative catch efficiency can depend...
on whether effects of length on dispersion are considered. The population size composition
can change through a variety of phenomena such as a pulse in recruitment, stock rebuilding,
or changes in selective fishing pressure. As such, I recommend smoothers to account for size
effects for relative catch efficiency and dispersion when analyzing paired haul data whenever
possible and the D-m model is a natural approach for these analyses. Although orthogonal
polynomials provided better fits with regard to AIC\textsubscript{c}, regression spline smoothers may be
useful in application to new data with lengths observed outside the range in the comparison
study.

**Acknowledgments**

I thank the Ecosystems Surveys Branch at the Northeast Fisheries Science Center as well
as the many volunteers that participated on board the *Albatross IV* and *Henry B. Bigelow*
during the calibration study. Many people helped with initial analyses of the calibration
experiments from which this work draws. Paul Rago had the initial idea to look at work by
others on binomial models that condition on the total catch from which this work germinated.
Chris Legault provided many constructive comments that greatly improved the quality of
this manuscript.
References


Byrne, C. J. and Forrester, J. 1991. Relative fishing power of NOAA R/Vs Albatross IV and Delaware II. NEFSC SAW/12/P1. Northeast Fisheries Science Center, Woods Hole, MA.


Hierarchical model for relative catch efficiency


Hierarchical model for relative catch efficiency


Hierarchical model for relative catch efficiency


Table 1. The (effective) degrees of freedom attributed to estimation of relative catch efficiency \( \rho \) and dispersion parameter \( \phi \), whether sampling fraction \( f \) and swept area \( A \) covariates were used in the \( \phi \) model, total degrees of freedom used by each fitted model, log-likelihood, and AIC\(_c\) for the ten best performing models of those considered for Acadian redfish as measured by \( \Delta(\text{AIC}_c) \). The performance of the best models in the OP-G and SP classes provided at the bottom of the table was relatively poor.

<table>
<thead>
<tr>
<th>Model type</th>
<th>( \rho ) df</th>
<th>( \phi ) df</th>
<th>( \phi ) covariates</th>
<th>Total df</th>
<th>( \log(\mathcal{L}) )</th>
<th>AIC(_c)</th>
<th>( \Delta(\text{AIC}_c) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>OP</td>
<td>8.00</td>
<td>5.00</td>
<td>( f, A )</td>
<td>13.00</td>
<td>-2506.31</td>
<td>5038.90</td>
<td>0.00</td>
</tr>
<tr>
<td>OP</td>
<td>8.00</td>
<td>4.00</td>
<td>( f )</td>
<td>12.00</td>
<td>-2507.51</td>
<td>5039.25</td>
<td>0.36</td>
</tr>
<tr>
<td>OP</td>
<td>9.00</td>
<td>5.00</td>
<td>( f, A )</td>
<td>14.00</td>
<td>-2505.95</td>
<td>5040.22</td>
<td>1.32</td>
</tr>
<tr>
<td>OP</td>
<td>8.00</td>
<td>6.00</td>
<td>( f, A )</td>
<td>14.00</td>
<td>-2506.08</td>
<td>5040.48</td>
<td>1.58</td>
</tr>
<tr>
<td>OP</td>
<td>9.00</td>
<td>4.00</td>
<td>( f )</td>
<td>13.00</td>
<td>-2507.14</td>
<td>5040.54</td>
<td>1.65</td>
</tr>
<tr>
<td>OP</td>
<td>8.00</td>
<td>5.00</td>
<td>( f )</td>
<td>13.00</td>
<td>-2507.23</td>
<td>5040.74</td>
<td>1.84</td>
</tr>
<tr>
<td>OP</td>
<td>9.00</td>
<td>6.00</td>
<td>( f, A )</td>
<td>15.00</td>
<td>-2505.69</td>
<td>5041.74</td>
<td>2.84</td>
</tr>
<tr>
<td>OP</td>
<td>8.00</td>
<td>4.00</td>
<td>( f, A )</td>
<td>12.00</td>
<td>-2508.80</td>
<td>5041.82</td>
<td>2.93</td>
</tr>
<tr>
<td>OP</td>
<td>10.00</td>
<td>5.00</td>
<td>( f, A )</td>
<td>15.00</td>
<td>-2505.77</td>
<td>5041.89</td>
<td>2.99</td>
</tr>
<tr>
<td>OP</td>
<td>9.00</td>
<td>5.00</td>
<td>( f )</td>
<td>14.00</td>
<td>-2506.82</td>
<td>5041.95</td>
<td>3.06</td>
</tr>
<tr>
<td>OP-G</td>
<td>8.00</td>
<td>2.00</td>
<td>( f, A )</td>
<td>10.00</td>
<td>-2557.51</td>
<td>5135.17</td>
<td>96.28</td>
</tr>
<tr>
<td>SP</td>
<td>9.33</td>
<td>5.44</td>
<td>( f, A )</td>
<td>14.77</td>
<td>-2507.57</td>
<td>5045.03</td>
<td>6.13</td>
</tr>
</tbody>
</table>
Table 2. Number of stations where fish were observed, number of those stations where a subsample was measured for length, total number of fish measured for length, minimum and maximum length (cm) and total number of observations by station and 1 cm size class for each species.

<table>
<thead>
<tr>
<th>Species</th>
<th># Stations</th>
<th># Subsamples</th>
<th># Fish</th>
<th>Minimum Length</th>
<th>Maximum Length</th>
<th># Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black sea bass</td>
<td>111</td>
<td>1</td>
<td>1203</td>
<td>3</td>
<td>61</td>
<td>511</td>
</tr>
<tr>
<td>Haddock</td>
<td>153</td>
<td>23</td>
<td>14618</td>
<td>6</td>
<td>77</td>
<td>2374</td>
</tr>
<tr>
<td>Atlantic cod</td>
<td>130</td>
<td>0</td>
<td>1797</td>
<td>4</td>
<td>121</td>
<td>1009</td>
</tr>
<tr>
<td>Summer flounder</td>
<td>176</td>
<td>0</td>
<td>1920</td>
<td>15</td>
<td>77</td>
<td>1226</td>
</tr>
<tr>
<td>Winter flounder</td>
<td>144</td>
<td>4</td>
<td>2959</td>
<td>6</td>
<td>61</td>
<td>1238</td>
</tr>
</tbody>
</table>
For species where the best OP-G and SP models are not among the best five models overall, the performance of the best models in those classes are also provided for comparison.

<table>
<thead>
<tr>
<th>Model Type</th>
<th>$\rho$ df</th>
<th>$\phi$ df</th>
<th>$\phi$ covariates</th>
<th>Total df</th>
<th>log($\mathcal{L}$)</th>
<th>$\text{AIC}_c$</th>
<th>$\Delta(\text{AIC}_c)$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Black Sea Bass</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OP-G</td>
<td>5.00</td>
<td>3.00</td>
<td>8.00</td>
<td>$-399.45$</td>
<td>815.19</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>OP-G</td>
<td>5.00</td>
<td>1.00</td>
<td>6.00</td>
<td>$-401.56$</td>
<td>815.28</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>OP</td>
<td>5.00</td>
<td>2.00</td>
<td>7.00</td>
<td>$-401.19$</td>
<td>816.48</td>
<td>1.29</td>
<td></td>
</tr>
<tr>
<td>OP-G</td>
<td>5.00</td>
<td>2.00</td>
<td>7.00</td>
<td>$-401.18$</td>
<td>816.58</td>
<td>1.39</td>
<td></td>
</tr>
<tr>
<td>OP</td>
<td>5.00</td>
<td>3.00</td>
<td>8.00</td>
<td>$-400.45$</td>
<td>817.04</td>
<td>1.85</td>
<td></td>
</tr>
<tr>
<td>SP</td>
<td>7.08</td>
<td>1.00</td>
<td>8.08</td>
<td>$-403.28$</td>
<td>823.02</td>
<td>7.83</td>
<td></td>
</tr>
<tr>
<td><strong>Haddock</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OP</td>
<td>11.00</td>
<td>7.00</td>
<td>$f,A$</td>
<td>18.00</td>
<td>$-3328.71$</td>
<td>6693.65</td>
<td>0.00</td>
</tr>
<tr>
<td>OP</td>
<td>12.00</td>
<td>7.00</td>
<td>$f,A$</td>
<td>19.00</td>
<td>$-3328.44$</td>
<td>6695.12</td>
<td>1.47</td>
</tr>
<tr>
<td>OP</td>
<td>11.00</td>
<td>8.00</td>
<td>$f,A$</td>
<td>19.00</td>
<td>$-3328.69$</td>
<td>6695.63</td>
<td>1.98</td>
</tr>
<tr>
<td>SP</td>
<td>5.01</td>
<td>10.41</td>
<td>$f,A$</td>
<td>15.42</td>
<td>$-3332.33$</td>
<td>6695.71</td>
<td>2.06</td>
</tr>
<tr>
<td>SP</td>
<td>5.04</td>
<td>9.41</td>
<td>$f$</td>
<td>14.45</td>
<td>$-3333.60$</td>
<td>6696.27</td>
<td>2.62</td>
</tr>
<tr>
<td>OP-G</td>
<td>11.00</td>
<td>5.00</td>
<td>16.00</td>
<td>$-3370.11$</td>
<td>6772.46</td>
<td>78.81</td>
<td></td>
</tr>
<tr>
<td><strong>Atlantic cod</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OP</td>
<td>9.00</td>
<td>3.00</td>
<td>$A$</td>
<td>13.00</td>
<td>$-770.86$</td>
<td>1567.89</td>
<td>0.00</td>
</tr>
<tr>
<td>OP</td>
<td>9.00</td>
<td>3.00</td>
<td></td>
<td>12.00</td>
<td>$-771.88$</td>
<td>1567.90</td>
<td>0.01</td>
</tr>
<tr>
<td>OP-G</td>
<td>9.00</td>
<td>3.00</td>
<td></td>
<td>12.00</td>
<td>$-772.70$</td>
<td>1569.71</td>
<td>1.82</td>
</tr>
<tr>
<td>OP</td>
<td>10.00</td>
<td>3.00</td>
<td></td>
<td>13.00</td>
<td>$-771.86$</td>
<td>1569.89</td>
<td>2.00</td>
</tr>
<tr>
<td>OP</td>
<td>10.00</td>
<td>3.00</td>
<td>$A$</td>
<td>14.00</td>
<td>$-770.85$</td>
<td>1569.89</td>
<td>2.00</td>
</tr>
<tr>
<td>SP</td>
<td>8.84</td>
<td>1.00</td>
<td></td>
<td>9.84</td>
<td>$-789.80$</td>
<td>1599.50</td>
<td>31.60</td>
</tr>
<tr>
<td><strong>Summer Flounder</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OP-G</td>
<td>5.00</td>
<td>2.00</td>
<td></td>
<td>7.00</td>
<td>$-957.81$</td>
<td>1929.72</td>
<td>0.00</td>
</tr>
<tr>
<td>OP-G</td>
<td>5.00</td>
<td>1.00</td>
<td></td>
<td>6.00</td>
<td>$-958.87$</td>
<td>1929.80</td>
<td>0.08</td>
</tr>
<tr>
<td>OP</td>
<td>5.00</td>
<td>1.00</td>
<td></td>
<td>6.00</td>
<td>$-959.18$</td>
<td>1930.40</td>
<td>0.68</td>
</tr>
<tr>
<td>SP</td>
<td>5.56</td>
<td>1.00</td>
<td></td>
<td>6.56</td>
<td>$-958.71$</td>
<td>1930.62</td>
<td>0.89</td>
</tr>
<tr>
<td>OP-G</td>
<td>5.00</td>
<td>4.00</td>
<td></td>
<td>9.00</td>
<td>$-956.54$</td>
<td>1931.24</td>
<td>1.51</td>
</tr>
<tr>
<td><strong>Winter Flounder</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OP-G</td>
<td>5.00</td>
<td>9.00</td>
<td></td>
<td>14.00</td>
<td>$-1059.68$</td>
<td>2147.70</td>
<td>0.00</td>
</tr>
<tr>
<td>OP</td>
<td>5.00</td>
<td>3.00</td>
<td>$f,A$</td>
<td>8.00</td>
<td>$-1065.98$</td>
<td>2148.04</td>
<td>0.34</td>
</tr>
<tr>
<td>OP-G</td>
<td>3.00</td>
<td>9.00</td>
<td></td>
<td>12.00</td>
<td>$-1061.95$</td>
<td>2148.16</td>
<td>0.46</td>
</tr>
<tr>
<td>SP</td>
<td>7.47</td>
<td>3.00</td>
<td>$f,A$</td>
<td>10.47</td>
<td>$-1063.58$</td>
<td>2148.30</td>
<td>0.60</td>
</tr>
<tr>
<td>OP</td>
<td>5.00</td>
<td>4.00</td>
<td>$f,A$</td>
<td>9.00</td>
<td>$-1065.11$</td>
<td>2148.32</td>
<td>0.63</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1. Of stations where swept area was observed for both gears, the number of stations where Acadian redfish were observed by the Albatross IV or Henry B. Bigelow only as well as by both gears (top), and the total number of fish observed (bottom) in tows by each gear by 1 cm length class.

Figure 2. Randomized quantile residuals of the best performing model (as measured by AICc, see Table 1) for Acadian redfish in relation to the predicted number captured by the Henry B. Bigelow (left), the total number of fish captured at a station (middle), and their normal quantiles (right).

Figure 3. Relative catch efficiency $\rho(L)$ of the Henry B. Bigelow to Albatross IV (top) and dispersion parameter $\phi(L)$ (bottom) as estimated for Acadian redfish from OP models where the degree of the polynomials for $\rho$ and $\phi$ are 7 and 2 (black lines, best performing model in Table 1) or 7 and 0 (gray lines), respectively, in comparison to the corresponding predictions from models fit to data in each 1 cm length class without any covariates (gray points). Vertical bars and dotted lines represent approximate 95% confidence intervals and the horizontal grey bar in the upper plot corresponds to equal catch efficiency ($\rho = 1$).

Figure 4. Of stations where swept area was observed for both gears, the number of stations where black sea bass, haddock, Atlantic cod, and summer and winter flounder were observed by the Albatross IV or Henry B. Bigelow only as well as by both gears (top), and the total number of fish observed (bottom) in tows by each gear by 2 cm length class.

Figure 5. Relative catch efficiency $\rho(L)$ of the Henry B. Bigelow to Albatross IV (top) and dispersion parameter $\phi(L)$ (bottom) as estimated by the best performing models (as measured by AICc, see Table 3) for black sea bass, haddock, Atlantic cod, summer flounder and winter flounder. Dotted lines represent approximate 95% confidence intervals and horizontal grey bars corresponds to equal catch efficiency ($\rho = 1$).
Hierarchical model for relative catch efficiency

Figure 1
Hierarchical model for relative catch efficiency
Figure 3
Figure 4

Hierarchical model for relative catch efficiency

- Black Sea Bass
- Haddock
- Atlantic cod
- Summer flounder
- Winter flounder

Length (cm)
Figure 5

Hierarchical model for relative catch efficiency

Black Sea Bass

Haddock

Atlantic cod

Summer Flounder

Winter Flounder

\( \rho(L) \)

\( \phi(L) \)

Length (cm)