Fisheries and Oceans Canada | Pêches et Océans Canada
Ecosystems and Oceans Science | Sciences des écosystèmes et des océans

Canadian Science Advisory Secretariat (CSAS)

Research Document 2022/002

Quebec Region

Re-analysis of comparative fishing experiments in the Gulf of St. Lawrence and other analyses to derive stock-wide bottom-trawl survey indices beginning in 1971 for 4RST Greenland halibut, *Reinhardtius hippoglossoides*

Yihao Yin[1] and Hugues P. Benoît[2]

[1]Fisheries and Oceans Canada
Bedford Institute of Oceanography
Dartmouth, NS B2Y 4A2

[2]Fisheries and Oceans Canada
Maurice Lamontagne Institute
Mont Joli, QC G5H 3Z4

June 2022

Canada

Foreword

This series documents the scientific basis for the evaluation of aquatic resources and ecosystems in Canada. As such, it addresses the issues of the day in the time frames required and the documents it contains are not intended as definitive statements on the subjects addressed but rather as progress reports on ongoing investigations.

Published by:

Fisheries and Oceans Canada
Canadian Science Advisory Secretariat
200 Kent Street
Ottawa ON K1A 0E6

http://www.dfo-mpo.gc.ca/csas-sccs/
csas-sccs@dfo-mpo.gc.ca

ISSN 1919-5044
ISBN 978-0-660-43823-8 Cat. No. Fs70-5/2022-002E-PDF

Correct citation for this publication:

Yin, Y. and Benoît, H.P. 2022. Re-analysis of comparative fishing experiments in the Gulf of St. Lawrence and other analyses to derive stock-wide bottom-trawl survey indices beginning in 1971 for 4RST Greenland halibut, *Reinhardtius hippoglossoides*. DFO Can. Sci. Advis. Sec. Res. Doc. 2022/002. viii + 45 p.

Aussi disponible en français :

Yin, Y. et Benoît, H.P. 2022. Réanalyse d'expériences de pêche comparative dans le golfe du Saint-Laurent et autres analyses visant à déterminer les indices de relevés au chalut de fond menés à l'échelle du stock de flétan du Groenland (Reinhardtius hippoglossoides) des divisions 4RST depuis 1971. Secr. can. des avis sci. du MPO. Doc. de rech. 2022/002. viii + 48 p.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Standardized bottom-trawl surveys provide fishery-independent estimates of relative abundance that are key to the assessment and management of demersal fish stocks worldwide. Over time it may be necessary or desirable to change the vessels, fishing gear or other protocols employed in the surveys to maintain or enhance survey efficiency. Such changes can affect the survey catchability of different species and sizes of fish that could otherwise be confounded with changes in stock abundance. Consequently comparative fishing experiments involving the old and new equipment and protocols are used to estimate calibration coefficients that are then applied to maintain the integrity of survey time series. Here we analyze data from previously analyzed and unanalyzed comparative fishing experiments to develop indices for Greenland halibut in NAFO 4RST that combine for the first time. data from surveys in the northern Gulf of St. Lawrence (GSL) and the southern GSL (sGSL), and which extend the survey series back to 1984. Currently, the main index used in the assessment for this stock begins in 1990 and is based on the northern GSL survey, which covers most but not all of the distribution of Greenland halibut in the GSL. We then show how the abundance index from the sGSL survey can be used to predict a GSL-wide index using a heuristic model of density-dependent distribution change. This approach is then applied to the sGSL index, available since 1971, to provide a longer-term perspective on the dynamics of the stock, particularly in light of fishery exploitation rates that were particularly elevated at times from the early 1970s to the mid-1990s. Although some refinements are required to improve the approach, these new indices provide important insights on the productivity, dynamics, and exploitation history for the stock.

1. INTRODUCTION

Standardized bottom-trawl surveys provide fishery-independent estimates of relative abundance that are key to the assessment and management of demersal fish stocks worldwide. By employing standardized and statistically-sound sampling schemes, and ideally covering large geographical areas that contain almost all of the stock's spatial distribution, these surveys minimize the impact of factors that may confound the interpretation of abundance trends, as is common for fishery-dependent indices (Hilborn and Walters 1992). In Atlantic Canada, annual research vessel bottom-trawl surveys have been in place since the early 1970s or 1980s, depending on the area (Chadwick et al. 2007; Benoît et al. 2020). Over this time-period it has periodically been necessary or desirable to change the vessels, fishing gear or other protocols employed in the surveys to maintain or enhance survey efficiency. Such changes can affect the survey catchability of different species and sizes of fish and macro-benthos that could otherwise be confounded with changes in stock abundance (e.g., Pelletier 1998; Lewy et al. 2004; Cadigan and Dowden 2010). Calibration experiments based on comparative fishing are the standard for estimating and accounting for changes in relative catchability caused by structural changes in the surveys (e.g., Miller 2013). In Atlantic Canada, these experiments have largely comprised concurrent side-by-side fishing of the new vessel/gear/protocol and that being replaced.

Two research vessel surveys are conducted annually in the Gulf of St. Lawrence (GSL), one covering the southern GSL (sGSL survey) and one covering the northern GSL and Estuary (hereafter, nGSL survey) (Figures 1 and 2).

The sGSL survey has been conducted annually each September since 1971, first by the E.E. Prince fishing a Yankee 36 trawl (1971–1985), followed by three vessels, each fishing a Western IIA trawl: the Lady Hammond (1985–1991), the Canadian Coast Guard Ship (CCGS) Alfred Needler (1992–2005) and the CCGS Teleost (2004–present). Comparative fishing experiments involving former and replacement vessels were conducted during the regular surveys in 1985, 2004 and 2005, and in a dedicated survey in August 1992, and the results of these experiments have been used to account for changes in catchability associated with the changes in vessels and trawl (Benoît and Swain 2003a; Benoît 2006). In addition, the sGSL survey changed from daylight-only to 24-hr operations in 1984, and effects on relative catchability of this change in protocol have been estimated using both comparative fishing day and night and at specific locations, and based on statistical modelling of survey catches (Benoît and Swain 2003b).

The nGSL survey has been conducted annually each August since 1984, first by the Lady Hammond fishing a Western IIA trawl (1984–1990), then by the CCGS Alfred Needler fishing a URI trawl (1990–2005) and the CCGS Teleost fishing a Campelen trawl (2004–present). Comparative fishing experiments involving former and replacement vessels and trawls were conducted during the regular surveys in 1990, 2004 and 2005. The results of the 2004–2005 experiments have been analyzed and the estimated calibration factors are routinely applied to maintain the integrity of the standardized abundance series for a large number of taxa (Bourdages et al. 2007). In contrast, there has been no extensive formal analysis of the data from the 1990 Lady Hammond and Alfred Needler comparison. In an unpublished working paper, Gascon et al. (1991) concluded that the results of that experiment were too variable to provide reliable conversion factors based on the analytical methods available at the time, although the basis for this conclusion is not clear from sparse results the authors presented. Nonetheless, accepted conversion factors have since been developed from these data for two species. Swain et al. (1998) developed conversion factors for witch flounder (*Glyptocephalus cynoglossus*). Conversion factors for redfish (*Sebastes* sp.) developed from these experiments

have been used to extend the survey series to 1984 for stock assessment (beginning with McAllister and Duplisea 2016, and Duplisea et al. 2016); however, the details and results of the analysis of the 1990 comparative fishing data have not been published. Lack of interest in analyzing the 1990 data to extend survey series back to 1984 for other species has likely resulted from Gascon et al.'s (1991) pessimistic conclusions and the fact that the nGSL survey undertaken by the Lady Hammond covered only deeper water strata (depths generally >100 m) that do not fully cover the summer distribution of species such as Atlantic cod.

It is well accepted that in the GSL, the abundance and productivity of numerous fishery resources, and ecosystem structure and functioning in general, changed considerably from the 1980s to the 1990s, and subsequently (Savenkoff et al. 2007a,b; Benoît and Swain 2008). The absence of survey indices for a majority of taxa in the nGSL for the 1980s greatly constrains our understanding of their productivity and their response to fishing and environmental change. For example, abundance and biomass of Greenland halibut increased considerably during the 1990s, although it is not clear to what extent this is due to favorable environmental conditions or stock recovery from overfishing caused by a spike in fishery landings in the latter half of the 1980s (Gauthier et al. 2020). As another example, it is clear that the collapse of the 3Pn, 4RS (nGSL) cod stock began several years before the start of the current survey time series in 1990 (Brassard et al. 2019). The inability to adequately estimate fishing and natural mortality rates prior to the collapse due to the absence of age-specific survey abundance estimates risks producing bias in the estimates of pre-collapse stock abundance and productivity, and in reference points used in the assessment and management of the stock.

The sophistication, flexibility and apparent robustness of statistical methods for the analysis of comparative fishing data have greatly improved over the past two decades (Pelletier 1998; Cadigan and Dowden 2010; Miller 2013). These methods are arguably better adapted than their predecessors to deal with variable data such as those from the 1990 Hammond-Needler trials.

The estimation and application of Hammond-Needler conversion factors for the nGSL survey presents advantages for several stocks in addition to simply prolonging the survey series by seven years. Neither the nGSL or sGSL surveys individually cover the entire distributional area for stocks for which the management area encompasses NAFO areas 4RST (e.g., 4RST Greenland halibut, redfish, witch flounder and Atlantic halibut), although joint survey coverage is (near) complete. With the exception of 4RST witch flounder (Ricard and Swain 2018), assessments for these stocks presently interpret abundance indices from each survey separately, which is at best sub-optimal. Because both the sGSL and nGSL surveys used the Lady Hammond fishing the Western IIA in some years, and calibrations for subsequent changes in vessel and gear exist or are possible, the development of a single standardized GSL survey series, such as that for witch flounder (Ricard and Swain 2018), is possible for many taxa, notably 4RST Greenland halibut and redfish in Unit 1. This approach is not viable for stocks such as 4RST Atlantic halibut for which survey catches were too small and infrequent in the 1980s and early 1990s to adequately estimate calibration factors allowing adjustment of data to a common vessel and gear based on comparative fishing data alone. However, for this stock and others, information on the relative catchability of the two surveys can be extracted by comparing their catches in an area of overlap along the southern slope of the Laurentian channel (e.g., Morin et al. 2016). This comparison can form the basis for standardizing the two surveys, or could eventually be combined with the results of comparative fishing in an integrated analysis (sensu Maunder and Punt 2013; Yin and Benoît 2022) to improve the precision of conversion factors used to produce a standardized series.

In this document we present for the first time the 4RST Greenland halibut results and analyses of the 1990 Lady Hammond and Alfred Needler comparative fishing experiment. We also re-analyse the data from the other existing comparative fishing experiments in the GSL using

modern statistical methods to derive a series of length-dependent conversions factors that can be applied to convert survey catches to a common standard, here catch by the CCGS Teleost using the Campelen trawl. We then use these standardized catches to derive whole-stock indices of abundance, biomass and exploitation rate for NAFO 4RST Greenland halibut.

2. METHODS

2.1. DATA

The data for this analysis all come from standardized survey sampling, mostly in the context of regular survey operations that in some cases also involved comparative fishing, but also from a dedicated comparative fishing experiment in the case of the August 1992 sGSL Hammond-Needler trials. A summary of survey vessels and survey protocols employed are provided in Table 1, and a summary of the survey trawls is provided in Table 2. Nominal target swept area, calculated as the product of tow distance (Table 1) and wingspread (Table 2), varied between vessels and trawls, as well as across years in the case of the first few years of the CCGS Alfred Needler in the nGSL.

Both surveys follow a stratified random survey design, with survey strata defined independently in each survey based on bathymetry and area (Figures 1 and 2; note that stratum numbering is survey-specific and there is no correspondence between similarly numbered strata in the two surveys). In the sGSL survey, strata 415–439 have been part of the design since 1971, and three coastal strata were added in 1984 (strata 401–403). In the nGSL survey, the strata retained in the sampling plan has varied over the years although a core group of strata has been sampled annually since at least 1985 (Figure 3). We draw the attention of readers to the strata that comprise the area of overlap between the two surveys: 415, 425, and 439 in the sGSL survey, and 401–406 in the nGSL survey (Figure 4). Catches in the area of overlap by the two surveys are those that provide information on relative catchability between concurrent survey vessels and gears. Further details on the surveys are available in Hurlbut and Clay (1990), Chadwick et al. (2007) and Bourdages et al. (2020).

Specific details for all but one comparative fishing experiments treated in our analyses are available in other reports and are not repeated here, with the exception of showing the location of the paired fishing sets (Figure 5; for additional details see Benoît and Swain 2003a; Benoît 2006; Bourdages et al. 2007). Because the details of the August 1990 nGSL Hammond-Needler trials have not been published, these are described briefly below.

Comparative fishing between the Lady Hammond with Western IIA trawl and the CCGS Alfred Needler with the URI trawl took place between August 22 and 29, 1990. During the paired tows, the vessels fished simultaneously and in parallel, separated by the shortest distance considered practical and safe. Standard fishing procedures proper to each vessel were employed (see Table 1). A tow on the Lady Hammond was considered initiated when the winches deploying the trawl warps were blocked and completed when the winches began hauling in the gear. In contrast, a tow on the CCGS Alfred Needler was considered initiated when parameters reported by trawl-mounted Scanmar sensors indicated the trawl was on bottom and adequately open, and completed when the trawl left the bottom. Effects on catchability of this and other differences in protocol (e.g., tow speed and duration; Table 1) should be accounted for in conversion factors estimated from comparative fishing data. A total of 80 valid paired tows were completed, at depths ranging from 74 to 486 m (average 267 m). The locations of the comparative fishing tows are shown in Figure 6.

2.2. COMPARATIVE FISHING DATA ANALYSIS

2.2.1. Binomial models

In the analysis of comparative fishing data, the goal is to estimate the relative fishing efficiency between a pair of vessels-gear combinations (referred to as gear in this section), denoted as *A* and *B*. We assume the expected catch from gear g ($g \in \{A, B\}$) at length l and at station i is

$$E[C_{gi}(l)] = q_{gi}(l)D_{gi}(l)f_{gi}.$$

Here, $q_{gi}(l)$ is the catchability of gear g at station i, D_{gi} is the underlying population density sampled by gear g, and f_{gi} is a standardization term for fishing effort which usually includes the swept area of a tow and if applicable, the proportion of sub-sampling for size measurement on-board. Station variation is considered in the catchability to account for possible differences in gear performance related to tow location. In a binomial model (e.g., Miller 2013), the catch from gear A at station i, conditioning on the combined catch from both gears at this station, $C_i(l) = C_{Ai}(l) + C_{Bi}(l)$, is binomial-distributed

$$C_{Ai}(l) \sim BI(C_i(l), p_{Ai}(l)),$$

where $p_{Ai}(l)$ is the expected proportion of catch from gear A. For paired tows, we usually assume the underlying densities at the station are the same, as the paired vessels are typically kept within a small distance of each other while fishing: $D_{Ai}(l) = D_{Bi}(l) = D_i(l)$. Then the logit-probability of catch by gear A is

$$logit(p_{Ai}(l)) = log\left(\frac{p_{Ai}(l)}{p_{Bi}(l)}\right) = log\left(\frac{E[C_{Ai}(l)]}{E[C_{Bi}(l)]}\right) = log(\rho_i(l)) + o_i.$$

$\rho_i(l)$ is the ratio of catchabilities between gear A and B at length l and at station i, or the conversion factor, which is the quantity of interest,

$$\rho_i(l) = q_{Ai}(l)/q_{Bi}(l),$$

and $o_i = log(f_{Ai}/f_{Bi})$ is an offset term derived from known standardization terms of the survey tows.

For a length-based conversion factor, we consider a smooth length effect based on a general additive smooth function,

$$log(\rho(l)) = \sum_{k=0}^{K} \beta_k X_k(l) = \mathbf{X}^T\boldsymbol{\beta},$$

where $\boldsymbol{\beta}$ are the coefficient parameters and are estimated, $\mathbf{X}$, or $\{X_k(l), k = 0,1,\cdots,K\}$, are a set of smoothing basis functions, and K is the dimension of the basis which controls the number of coefficient parameters and is usually pre-defined. In this study, we used the cubic spline smoother (Hastie et al. 2009), and the basis functions and penalty matrices were generated by the R package.

The estimation of a cubic spline smoother is based on the penalized sum of squares smoothing objective, but in practice, this is usually replaced by a penalized likelihood objective (Green and Silverman 1993):

$$\mathcal{L}(\boldsymbol{\beta}, \lambda) = f(\mathbf{Y}|\mathbf{X}, \boldsymbol{\beta})e^{-\frac{\lambda}{2}\boldsymbol{\beta}^T\mathbf{S}\boldsymbol{\beta}},$$

Where $\mathcal{L}$ denotes the likelihood objective function. $f(\mathbf{Y}|\mathbf{X}, \boldsymbol{\beta})$ is the joint probability function of the survey data $\mathbf{Y}$ conditional on the basis functions and coefficient parameters. $\mathbf{S}$ is the penalty matrix defined by the smoother and the dimension of the basis, and λ is the smoothness

parameter. This smoothness parameter can be estimated by maximum likelihood along with other model parameters but may be sensitive to the data. In such cases, it can be determined by other criteria such as generalized cross-validation (Wood 2000).

The penalized maximum likelihood smoother can also be re-parameterized into a mixed effects model (Verbyla et al. 1999; Wood 2017) to facilitate implementation as well as incorporation of additional random effects:

$$log(\rho(l)) = \mathbf{X}_f^T \boldsymbol{\beta}_f + \mathbf{X}_r^T \mathbf{b},$$

where $\boldsymbol{\beta}_f$ are fixed effects and $\mathbf{b}$ are random effects. $\mathbf{X}_f$ and $\mathbf{X}_r$ are transformed from the basis functions $\mathbf{X}$ and an eigen-decomposition of the penalty matrix $\mathbf{S}$, $\mathbf{X}_f = \mathbf{U}_f^T \mathbf{X}$ and $\mathbf{X}_r = \mathbf{U}_r^T \mathbf{X}$, where $\mathbf{U}_f$ and $\mathbf{U}_r$ are the eigenvectors that correspond to the zero and positive eigenvalues of $\mathbf{S}$. The random effects $b \sim \mathrm{N}(0, \mathbf{D}_+^{-1}/\boldsymbol{\lambda})$ where D_+ is the diagonal matrix of the positive eigenvalues of S. In the mixed effects model representation of the cubic spline smoother, the number of fixed effects is 2 and the number of random effects is therefore $K - 2$. Smoothing effects are transformed into shrinkage of random effects in the fitting of random deviations, and can be integrated into complex mixed effects models commonly used in fisheries science (Thorson and Minto 2015).

Additional random effects can be incorporated into the mixed effects model to address variations in the relative catch efficiency related to each station,

$$log\big(\rho_i(l)\big) = \mathbf{X}_f^T \big(\boldsymbol{\beta}_f + \boldsymbol{\delta}_i\big) + \mathbf{X}_r^T (\mathbf{b} + \boldsymbol{\epsilon}_i),$$

where $\boldsymbol{\delta}_i \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$ and $\boldsymbol{\epsilon}_i \sim \mathrm{N}(\mathbf{0}, \mathbf{D}_+^{-1}/\xi)$. When station variation is considered, the cubic smoothing over length applies to $\rho_i(l)$ at each station i. The random variables $\boldsymbol{\delta}_i$ and $\boldsymbol{\epsilon}_i$ allow for deviations of the length-based conversion at each station and are derived from a similar re-parameterization of the cubic spline smoother. For cubic smoothing, $\boldsymbol{\Sigma}$ is a two-dimensional, diagonal covariance matrix of the random effects corresponding to the random deviations thus containing three parameters. ξ controls the degree of smoothness of the random smoothers and the smoother at each station can differ.

A summary of the above binomial mixed model is as follows,

$$C_i(l) = C_{Ai}(l) + C_{Bi}(l),$$

$$C_{Ai}(l) \sim BI(C_i(l), p_{Ai}(l)),$$

$$logit(p_{Ai}(l)) = log(\rho_i(l)) + o_i,$$

$$log(\rho_i(l)) = \mathbf{X}_f^T (\boldsymbol{\beta}_f + \boldsymbol{\delta}_i) + \mathbf{X}_r^T (\mathbf{b} + \boldsymbol{\epsilon}_i).$$

The model is estimated via maximum likelihood estimation using all observations ($i = 1, 2, \cdots, m$) and the marginal likelihood integrating out random effects is

$$\mathcal{L}(\boldsymbol{\beta}_f, \boldsymbol{\Sigma}, \lambda, \xi) = \int \Big(\prod_{i=1}^{m} \int \int f(\mathbf{Y}_i | \mathbf{X}_f, \mathbf{X}_r, \boldsymbol{\beta}_f, \mathbf{b}, \boldsymbol{\delta}_i, \boldsymbol{\epsilon}_i) f(\boldsymbol{\delta}_i | \boldsymbol{\Sigma}) f(\boldsymbol{\epsilon}_i | \xi) \mathrm{d}\boldsymbol{\delta}_i \mathrm{d}\boldsymbol{\epsilon}_i \Big) f(\mathbf{b} | \lambda) \mathrm{d}\mathbf{b}.$$

The binomial mixed model can be adapted for various assumptions on the smoother and potential station variation to accommodate different underlying density of a species and data limitations especially in length measurements. A set of binomial models considered in the present analyses is provided in Table 3.

2.2.2. Beta-binomial models

The binomial assumption for the catch can be extended to a beta-binomial distribution to explain over-dispersion at the stations (Miller 2013):

$$C_{A,i}(l) \sim BB(C_i(l), p_{A,i}(l), \phi_i(l)).$$

The beta-binomial distribution is a compound of the binomial distribution and a beta distribution. The beta distribution amounts to a "prior" distribution for the expected proportion of catch from gear A across stations to accommodate for increased variance in the catch. More specifically, the expected catch by gear A has a variance of

$$var(C_{A,i}) = C_i p_i (1 - p_i) \frac{\phi_i + C_i}{\phi_i + 1},$$

where ϕ is the over-dispersion parameter that captures the extra-binomial variation.

The same smoothed length effect can be applied to the over-dispersion parameter,

$$log(\phi_i(l)) = \mathbf{X}_f^T \boldsymbol{\gamma} + \mathbf{X}_r^T \mathbf{g},$$

where $\boldsymbol{\gamma}$ are fixed effects and $\mathbf{g}$ are random effects, $\mathbf{g} \sim \mathrm{N}(0, \mathbf{D}_+^{-1}/\tau)$. This length effect models the variance heterogeneity and is particularly useful for projecting uncertainty. However, estimation of a length-based variance parameter typically requires sufficient catch at length data, which is usually not available for less abundant species.

A summary of the beta-binomial mixed model is as follows,

$$C_i(l) = C_{Ai}(l) + C_{Bi}(l),$$
$$C_{Ai}(l) \sim BB(C_i(l), p_{Ai}(l), \phi_i(l)),$$
$$logit(p_{Ai}(l)) = log(\rho_i(l)) + o_i,$$
$$log(\rho_i(l)) = \mathbf{X}_f^T(\boldsymbol{\beta}_f + \boldsymbol{\delta}_i) + \mathbf{X}_r^T(\mathbf{b} + \boldsymbol{\epsilon}_i),$$
$$log(\phi_i(l)) = \mathbf{X}_f^T \boldsymbol{\gamma} + \mathbf{X}_r^T \mathbf{g}.$$

The marginal likelihood is

$$\mathcal{L}(\boldsymbol{\beta}_f, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \lambda, \xi, \tau) =$$
$$\int\int \left(\prod_{i=1}^{m} \int\int f(\mathbf{Y}_i|\mathbf{X}_f, \mathbf{X}_r, \boldsymbol{\beta}_f, \mathbf{b}, \boldsymbol{\gamma}, \mathbf{g}, \boldsymbol{\delta}_i, \boldsymbol{\epsilon}_i) f(\boldsymbol{\delta}_i|\boldsymbol{\Sigma}) f(\boldsymbol{\epsilon}_i|\xi) \mathrm{d}\boldsymbol{\delta}_i \mathrm{d}\boldsymbol{\epsilon}_i\right) f(\mathbf{b}|\lambda) f(\mathbf{g}|\tau) \mathrm{d}\mathbf{b} \mathrm{d}\mathbf{g}.$$

Likewise, various smoothing assumptions can be applied to the variance parameter. Table 4 presents a set of beta-binomial mixed models.

2.2.3. Model fitting and selection

In the analysis of each comparative fishing experiment, the conversion factor was developed for a pre-specified length range at an interval of 1 cm. The length range was usually selected as from the minimum to the maximum observed length from the comparative fishing survey. Extremely large or small individuals of sporadic catches may be excluded from the analysis to avoid disproportional impact from these high leverage cases.

The binomial and beta-binomial models in Tables 3 and 4 were implemented in *TMB* (Kristensen et al 2016) which were compiled into objective functions and subsequently optimized in *R*. The basis functions for the cubic smoothing spline and the corresponding penalty matrices were generated using the R package mgcv (Wood 2011) based on 10 knots ($K = 9$) equally-spaced within the pre-specified length range in each comparative fishing analysis. This value was that used in the original analyses by Miller (2013) and appeared reasonable given that length measurements generally ranged between 6 to 72 cm, depending on the survey. Nonetheless, we tested different values for K by comparing model estimates and AIC values and found there to be little sensitivity of results to this choice. In addition to the

maximum likelihood estimation of the conversion factor, TMB automatically calculates a standard error using the delta method (Kristensen et al. 2016) to inform prediction uncertainty.

There were in total 13 candidate models for estimating the conversion factors. The best model for each species and each comparative fishing survey was selected by AIC (Akaike information criterion) to maximize model fitting, while avoiding over-fitting of more complicated models especially in cases without adequate data. In each analysis, the estimated μ (expected proportion of catch by gear A) from all converged models were compared along with the sample proportions (aggregated by stations and averaged for each length) to provide a more rigorous interpretation of the results. The estimated ρ (expected relative catch efficiency, or conversion factor) from the best model is presented here and validated with estimation from other studies, when available.

2.2.4. Calibration of survey catch

The conversion factors estimated from the comparative fishing experiments were applied to the annual bottom trawl survey catches from the vessels E.E. Prince, MV Lady Hammond and the CCGS Alfred Needler in the respective survey areas to calibrate to catches equivalent to those that would be made by the CCGS Teleost fishing the Campelen trawl, C_{TC}. For most surveys this involved applying sequential length dependent conversion factors:

$C_{TC} = \rho(l)_{NU \rightarrow TC}\, C_{NU},$ for catches by the Needler fishing the URI trawl in the nGSL survey;

$C_{TC} = \rho(l)_{NU \rightarrow TC}\, \rho(l)_{LW \rightarrow NU}\, C_{LW},$ for catches by the Lady Hammond fishing the WIIA trawl;

$C_{TC} = \rho(l)_{NU \rightarrow TC}\, \rho(l)_{LW \rightarrow NU}\, \rho(l)_{LW \rightarrow PY}\, C_{PY},$ for catches by the Prince fishing the Yankee 36 trawl in the sGSL survey;

$C_{TC} = \rho(l)_{NU \rightarrow TC}\, \rho(l)_{LW \rightarrow NU}\, \rho(l)_{LW \rightarrow NW}\, C_{NW},$ for catches by the Needler fishing the WIIA trawl in the sGSL survey; and finally,

$C_{TC} = \rho(l)_{NU \rightarrow TC}\, \rho(l)_{LW \rightarrow NU}\, \rho(l)_{LW \rightarrow NW}\, \rho(l)_{NW \rightarrow TW}\, C_{TW},$ for catches by the Teleost fishing the WIIA trawl in the sGSL survey.

In this report we proceed with calibrations as they have traditionally been employed in time series produced by DFO, that is without propagating their uncertainty to the estimated uncertainty in catch related estimates, such as abundance indices. Propagation of uncertainty is reasonably straightforward in an integrated survey analysis model, or using computer intensive approaches such as bootstrapping, and is planned for the analysis of these survey data in the future.

To verify that the sequential application of calibration factors results in an adequate calibration, we compared the annual calibrated survey catches in both the nGSL and sGSL surveys in their area of overlap (Figure 4). Provided that estimated calibration factors are accurate, that the densities of fish in the area do not differ between surveys despite being conducted sequentially, and given that catches from all surveys are converted to a common vessel and gear, catches by the nGSL and sGSL surveys in the area of overlap should have similar magnitude every year.

Length-dependent relative catch efficiency was estimated only over the range of length available in the respective comparative fishing experiments. When applying these estimates to lengths below or above this range to calibrate survey catches, we assumed constant efficiencies equal, respectively, to those at the minimum and maximum lengths of the range in the estimation.

2.3. CALIBRATED SURVEY DATA ANALYSIS

Analyses were based on data from the surveys in the nGSL for 1984–2020 and the sGSL for 1971–2020. First we constructed abundance indices for the common 1984–2020 period. Then, we quantified the relationship between indices for the sGSL survey and the whole GSL for the common period, and applied that relationship to indices from the sGSL for 1971–1983 to infer whole GSL estimates for that period. Finally, we consider those indices in light of landings in the fishery to derive estimates of an annual exploitation rate.

Combining the two surveys to derive abundance indices required adopting a common stratification scheme. This was accomplished most efficiently by replacing strata 415, 425 and 439 in the sGSL with strata 401–406 in the nGSL (Figure 4). This removed any overlap between the two surveys, resulting in a continuous stratification scheme that covered the entire Gulf survey area. Sets in the sGSL survey in the area of overlap were assigned to strata 401–406 based on their geographic position. A small number of sGSL sets occurring in the southeast portion of stratum 439 not covered by the nGSL survey were dropped from the analyses (Figure 7). A lack of overlap between the two surveys in the northwestern portion of nGSL strata 403 and 406 results in uneven set densities in that area compared to the other part of those strata (Figure 7). This heterogeneity was ignored in the present analyses because the surface area in question is small relative to the GSL area and Greenland halibut density was assumed homogenous within strata, however future analyses should validate whether these assumptions are valid.

All surveys except those by the E.E. Prince operated 24 hours per day. However, previous analyses found that catchability of Greenland halibut to the sGSL surveys did not vary over the diel cycle (Benoît and Swain 2003b), consequently no adjustments were required for the inclusion of the E.E. Prince survey. The 2003 sGSL survey was undertaken by the CCGS Wilfred Templeman, the sister ship to the CCGS Alfred Needler, and their catchability was assumed equal for our analyses.

Catches were standardized to a common Teleost-Campellen equivalent and were adjusted for area. For analyses that combined data from both surveys, hauls in each of the overlapping strata were assumed to have resulted from random sampling, irrespective of survey, given that original selection was random by survey. Of course this assumption is incorrect for the northwestern portion of strata 403 and 406 sampled only by the nGSL survey (Figure 4); however, the impacts on the estimated abundance index are expected to be minor given the small surface area involved.

All analyses were based on a common set of strata sampled every year, or almost, in each survey. Specifically the analyses retained strata 415 to 439 in the sGSL survey and strata 401–414, 801–824 and 827–832 in the nGSL survey. Strata 401–403 from the sGSL survey were excluded because Greenland halibut do not occur there and these strata are excluded from sGSL specific time series for this species that begin in 1971 (Gauthier et al. 2020). Strata 835-841 from the nGSL survey were excluded because they only began being sampled in 1990 (Figure 3). Densities of Greenland halibut in these strata and stratum 433 are typically very small and exclusion of these strata to produce a single index for the entire GSL should not result in bias.

In both surveys, strata are occasionally left unsampled in a given year due to operational constraints (e.g., Figure 3). To ensure that abundance indices were proper, the mean and variance of catches in unsampled stratum-year combinations were imputed using linear models of log(catch+1) as a function of fixed effects for stratum and year (both as factors), based on catches in the survey in the year in question and the three preceding it. The only exception was

for unsampled strata in 1984–1987, where a five-year data window was employed to ensure sufficient data for imputation.

Relative abundance and biomass indices, with 95% confidence intervals, were estimated using the common estimators for stratified random samples. Abundance indices were estimated for size-aggregated catch, and for the size classes used in the assessment of the stock. Biomass indices were estimated based on the estimated weight of individual Greenland halibut derived from sex-specific annual length-weight relationships from the nGSL survey.

Analyses revealed that Greenland halibut have shown density dependent expansion into the northern portion of the sGSL when abundance increased, and retreat as it decreased. This results in an asymptotic relationship between the whole GSL indices of abundance and those for the sGSL. This relationship was modelled using a function analogous to the Beverton-Holt model, specifically:

$$I_{g,t}^{GSL} = \frac{\alpha\, I_{g,t}^{sGSL}}{\beta + I_{g,t}^{sGSL}},$$

where $I_{g,t}^{GSL}$ is the whole GSL abundance index for size group *g* in year *t*, $I_{g,t}^{sGSL}$ is the index for the sGSL, and α and β are parameters to be estimated. Parameters were estimated separately for the four size groups commonly used in the assessment of 4RST Greenland halibut (Gauthier et al. 2020): [0, 20] cm;]20, 30] cm;]30, 40] cm; and >40 cm. The model was fitted using non-linear least squares, via the nls() function in R. The model and parameters were then used to predict whole GSL indices for the 1971–1983 period, $\hat{I}_{g,t}^{GSL}$, based on estimates of $I_{g,t}^{sGSL}$. To evaluate model fit, relative residuals were calculated as ($\hat{I}_{g,t}^{GSL}$-$I_{g,t}^{sGSL}$)/ $\hat{I}_{g,t}^{GSL}$.

We also estimated a relative exploitation rate by dividing 4RST fishery landings by the estimated trawlable biomass of Greenland halibut >35 cm. Trawlable biomass was estimated using annual length-weight relationships and length specific catch rates in the surveys. Normally the exploitation rate used in the assessment for this stock is based on individuals >40 cm; however, the length range was extended slightly here because the fishery prior to 1990 captured smaller individuals compared to the later years (Gauthier et al. 2020).

3. RESULTS AND DISCUSSION

3.1. RELATIVE CATCH EFFICIENCY

For each of the five comparative fishing analyses, 13 candidate binomial and beta-binomial models were fitted to the data. Model convergence was rigorously checked by the maximum gradient and Hessian matrix in each model fitting to ensure proper convergence. The nGSL 2004–2005 comparative fishing survey had sufficient data for more complicated models while data from the other surveys typically only fit more simple models successfully. Model AICs were computed whenever there was proper convergence. Table 5 presents the difference in AIC from the model giving lowest AIC. The best model for each analysis was selected by the lowest AIC for the nGSL 1990, nGSL 2004–2005, sGSL 1992 and sGSL 2004–2005 comparative fishing analyses from all converged models. In the analysis of sGSL 1985 comparative fishing experiment, the candidate models were reduced to the three models without a length effect, i.e. BI0, BI1 and BB0, and the best model was subsequently selected by the lowest AIC among these three models. The models with a length effect were excluded from the selection because this survey did not result in sufficient data for accurate estimation of the length effect. More specifically, the number of nontrivial observations (a pair of tows with at least one catch) is 1.9 on average over length bins. In addition, the estimated conversion factors from these more

complicated models such as BI2, BI3 and BB2 indicated a lack of evidence for a significant length effect, despite their convergence and marginal improvements in AIC from BB0.

For each analysis, the expected proportion of catch from the old vessel $\mu(L)$ from each converged model was compared to the sample proportion of catch (computed for each station and for an average across stations) in Figures 8-12 to assess the effectiveness of different model assumptions.

The length effect was not significant or was subtle in all pairs except for nGSL 1990 between Alfred Needler-URI and Lady Hammond-WIIA (Figure 8). As a result, this length dependency of the relative catch efficiency between URI and WIIA estimated from nGSL 1990 provided information for the calibration of all other sGSL catches from WIIA.

A random station effect typically improved model fit, except for nGSL 1990, where BI2 without a station effect performed slightly better than BI3 with a station effect, but estimation results were effectively the same (Figure 8). A conversion factor developed by Bourdages et al. (2007) based on a parametric exponential model (length effect as an exponential function) was included in the results comparison of the nGSL 2004–2005 comparative fishing analysis. Results from the smoothing length effect differed from the exponential length effect for smaller sizes (< 20cm) and the inclusion of a station effect in BB5 indicated a slightly lower conversion factor (Figure 9) as results from the exponential model were close to BI2.

In order to assess model fits, prediction residuals were derived from the best models in each comparative fishing analysis. Residual diagnostics using normalized randomized quantile residuals (Dunn and Smyth 1996) did not show significant deviations related to either length or station (Figures 13-17). Residuals from sGSL 1985 (Figure 15) and 1992 (Figure 16) indicated relatively higher prediction bias for a few length bins, and higher prediction uncertainty overall likely mostly due to small sample size.

The estimated relative catch efficiency (conversion factor) between each pair of vessel-gears was estimated from the best selected model (Figure 18), and subsequently calibrated to Teleost-Campelen by sequential multiplication (Figure 19). These were then applied to respective survey catches to convert to Teleost-Campelen equivalents (calibrated catch). Figure 20 shows the survey-aggregated annual catches at length by each vessel, in each survey, for the area of survey overlap. This figure indicates that although length-dependent patterns in calibrated catches matched well between the two surveys, the sGSL surveys tended to catch more Greenland halibut in most years, including during the 1980s when the two surveys employed the same vessel and gear. In years when there are differences between the vessels, the magnitude of the difference is similar across years. Superficially this suggests that catchability or density of Greenland halibut differs between the sGSL and nGSL surveys due to factors other than the changes in vessels and gear, or differences in swept area, which were accounted for in the analysis. In other words, the calibration appear to correctly adjust the catches for vessel, gear and protocol related changes. There are two principal factors that differ between the surveys that could explain the differences in catches, namely the months in which they take place and station locations.

On average, there is a one-month difference in the timing of sampling of stations by the two surveys in the area of overlap. Seasonal movements into the area of overlap, either downslope from the shallower areas of the sGSL or upslope from the Laurentian channel could result in an increase in catch in the sGSL survey. This mechanism appears to be a plausible explanation for the differences observed. Catch rates in the area of overlap, particularly the northwestern portion with latitude > 48.0 N tend to be greater in the sGSL survey in most years, particularly for fish > 30 cm (Figure 21). This suggests an increase in local density from August to September; however, it is not possible to determine the provenance of these additional fish

(either from shallower or deeper areas) with these survey data, although this might be possible by comparing catches in the sGSL September survey with those in the sGSL sentinel mobile gear survey, which has taken place annually in August since 2003 (Savoie 2016). Identifying the source of these fish could improve the accuracy of abundance indices that combine the nGSL and sGSL because if the fish result from upslope movements, then some double counting will occur, producing a likely small positive bias, while downslope movements may be less problematic if fish are only moving with the sGSL area and not moving outside that area and therefore not being counted.

Sampling by the two surveys in the area of overlap is not homogenous. Notably, the sGSL generally samples a more restricted and temporally variable set of depths compared to the nGSL survey (Figure 22). Additionally, the sGSL survey tends to have more sets in the northwest corner of the area of overlap (latitude >49.0 N) where catches are greater (Figure 21), likely contributing to the differences in mean catches observed between the surveys.

Although the difference in catches between the nGSL and sGSL surveys in the area of overlap may indicate a possible bias for a combined index, its magnitude is likely to be small given that this area constitutes a small portion of the overall combined survey areas and of the depths favoured by Greenland halibut in the GSL, approximately 200 to 400 m (Figure 2). Furthermore, given that the difference in mean catch rates appeared roughly similar across most years (Figure 20), it may be that any bias is stationary over time, which would pose no problem for the abundance indices given they are treated as relative, i.e., the bias would effectively be confounded with catchability to the surveys.

3.2. CALIBRATED SURVEY DATA ANALYSIS

Size-class specific indices of abundance (Figure 23) and biomass (Figure 24), and length aggregated survey indices (Figure 25) were nearly identical to those presented previously based on the nGSL survey alone for the common 1990-2018 period (Gauthier et al. 2020). Although the scale of the indices is lower with the addition of the sGSL survey, where a large proportion of sets catch no Greenland halibut, the interannual patterns differ very little with those of the nGSL survey only. Abundance indices for the sGSL survey are presented and discussed later in this section.

Abundance and biomass of individuals 0-20 cm was low and decreasing from 1984-1988, then increased to a small peak before dropping again in 1993 (Figure 25). They then increased to a much higher average level about which they have fluctuated since. Abundance of individuals >20 to 30 cm generally lagged those of the smaller class by 1 year, beginning in 1984 at a relatively elevated, albeit highly uncertain level. Abundance of individuals >30 cm to 40 cm showed a first peak around 1985–1986 and again around 1992, before rising steeply with incoming recruitment. Following a peak in the early 2000s, the index declined consistently, perhaps leveling off in the last five years. Abundance of individuals >40 cm peaked in 1986, rose again around 1995 before increasing to a high level in 2004. The index has since declined over the past decade. Abundance by size class in the nGSL has previously been reported to demonstrate clear cohort dynamics for the post 1990 period (Gauthier et al. 2020). The new results presented here confirm that this pattern also occurred prior to 1990.

The size aggregated mean number per tow declined from 1984 to 1989, and displayed a small peak in the early 1990s before rising to an elevated level around 2000 (Figure 25, top panel). The index has fluctuated largely without trend since the mid-2000.

Trawlable suvey biomass displayed a small peak around 1987, and increased to a high level beginning in the late 1990s (Figure 25, bottom panel). After fluctuating without major trends for much of the 2000s, biomass declined consistently over the 2010s, increasing slightly in 2020.

During the late 1980s, size-aggregated catches of Greenland halibut were elevated only in the St. Lawrence estuary (Figure 26). Catches were progressively smaller in the eastward direction. Over time, catch magnitude increased particularly in the Laurentian, Anticosti and Esquiman channels. Beginning in the late 1990s both the occurrence and biomass of Greenland halibut increased in Baie des Chaleurs and the Cape Breton Trough. Catches also spread onto parts of the Magdalen shallows, although these comprised mainly smaller individuals (results not shown). Recent declines in biomass catch rates have occurred mainly in the eastern portion of the Gulf. These distributional patterns appear to reflect density dependent expansion and the retreat from a core area situated in the St. Lawrence estuary.

The relationship between whole GSL and sGSL only indices is consistent with density-dependent distributional change. The asymptotic model fit this relationship reasonably well in absolute terms, particularly for the two larger size groups and the sGSL index was thus a reasonable predictor of the whole GSL index (Figure 27). However, relative residuals for all size groups tended to be larger and positive for the early portion of the series when abundance was lower. Different density-dependent model formulations, possibly including a temporal random walk for parameters, should be explored in the future before using the predicted series in the assessment for this stock. Nonetheless, the predicted indices for the 1971–1983 period display the aforementioned cohort dynamics, characterized by the appearance of strong cohorts in the mid 1980s, early 1990s, and late 1990s through much of the 2000s. They also suggest that peaks in the abundance of larger Greenland halibut in the late 1970s, mid to late 1980s and then the 2000s were characterized by increasing densities over time leading to an increase in stock abundance.

There were important peaks in fishery landings in the late 1970s and late 1980s. These were associated with large increases in relative exploitation rate for individuals >35 cm (Figure 28). The peak in exploitation rate in the 1980s, for which whole GSL trawlable biomass indices were directly available, was around 25%. The same peak based on inferred trawlable biomass from the sGSL survey index and the density-dependent relationship was much higher (around 65%), indicating that inferred biomass was underestimated. It is therefore likely that inferred exploitation rates prior to 1984, which were very high at times, were also overestimated by this approach, which is a consequence of model underprediction of abundance prior to the early 1990s (Figure 27). Nonetheless, the results suggest that exploitation rates prior to the 1990s were much higher than in the subsequent period. They also indicate that the fishery rapidly increased removals each time the abundance of Greenland halibut > 30 cm increased, rapidly reducing these abundances to very low levels (Figures 27 and 28). In each case, increases in the abundance of larger Greenland halibut appears to have resulted from single large year classes (Figure 27).

Exploitation rates dropped below 5% for the first time in the mid 1990s and again around 2000 (Figure 28). These lower rates, combined with strong year classes born in the early 1990s and in many years of the 2000s, appear to be principal factors in the high levels of large fish abundance in the 2000s (Figure 27). Overall, these dynamics suggest that 4RST Greenland halibut were likely recruitment overfished during the 1980s and early 1990s. Although favorable environmental conditions may also have contributed to increased recruitment in the latter portion of the 1990s, recovery from fishery-related impairment may have been a much more important contributor than previously believed. Size-based analytical population modelling of this stock would help to elucidate the relative roles of fishing and environmental change on recruitment.

Basic density-dependent modelling of survey indices presented here suggests there is strong potential to develop abundance indices for 4RST Greenland halibut that begin in 1971. Further work is required to refine the use of sGSL survey information to inform whole-stock indices and

to develop confidence intervals that incorporate both survey sampling error and prediction error. Notably, the exploitation rate analysis indicates that whole GSL abundance is likely underestimated by the current density-dependent relationship with the sGSL index and therefore efforts to improve those predictions, which occur at low abundance, could reduce this bias.

4. ACKNOWLEDGEMENTS

We wish to thank Noel Cadigan and Jean-Martin Chamberland for kindly agreeing to review the penultimate draft of this document. We also thank Yoland Plourde for his help in translating certain terms in Table 2.

5. REFERENCES CITED

Benoît, H.P. 2006. Standardizing the southern Gulf of St. Lawrence bottom trawl survey time series: Results of the 2004-2005 comparative fishing experiments and other recommendations for the analysis of the survey data. DFO Can. Sci. Adv. Sec. Res. Doc. 2006/008.

Benoît, H.P., and Swain, D.P. 2003a. Standardizing the southern Gulf of St. Lawrence bottom-trawl survey time series: adjusting for changes in research vessel, gear and survey protocol. Can. Tech. Rep. Fish. Aquat. Sci. no. 2505: iv + 95 pp.

Benoît, H.P., and Swain, D.P. 2003b. 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: 1297-1316.

Benoît, H.P., and Swain, D.P. 2008. Impacts of environmental change and direct and indirect harvesting effects on the dynamics of a marine fish community. Can. J. Fish. Aquat. Sci. 65: 2088-2104.

Benoît, H.P., Dunham, A., Macnab, P., Rideout, R., Wareham, V., Clark, D., Duprey, N., Maldemay, É.-P., Richard, M., Clark, C., and Wilson, B. 2020. Elements of a framework to support decisions on authorizing scientific surveys with bottom contacting gears in protected areas with defined benthic conservation objectives. DFO Can. Sci. Advis. Sec. Res. Doc. 2020/011. ix + 98 p.

Bourdages, H., Savard, L., Archambault, D., and Valois, S. 2007. Results from the August 2004and 2005 comparative fishing experiments in the northern Gulf of St Lawrence between the CCGS *Alfred Needler* and the CCGS *Teleost*. Can. Tech. Rep. Fish. Aquat. Sci. 2750: ix + 57 pp.

Bourdages, H., Brassard, C., Desgagnés, M., Galbraith, P., Gauthier, J., Nozères, C., Scallon-Chouinard, P.-M. and Senay, C. 2020. Preliminary results from the ecosystemic survey in August 2019 in the Estuary and northern Gulf of St. Lawrence. DFO Can. Sci. Advis. Sec. Res. Doc. 2020/009. iv + 93 p.

Brassard, C., Lussier, J-F., Benoît, H, Way, M. and Collier, F. 2019. The status of the Northern Gulf of St. Lawrence (3Pn, 4RS) Atlantic cod (*Gadus morhua*) stock in 2018. DFO Can. Sci. Advis. Sec. Res. Doc. 2019/075. xi + 117 p.

Cadigan, N.G., and Dowden, J.J. 2010. Statistical inference about the relative efficiency of a new survey protocol, based on paired-tow survey calibration data. Fish. Bull. 108: 15-29.

Chadwick, E.M.P., Brodie, W., Clark, D., Gascon, D., and Hurlbut, T.R. 2007. History of annual multi-species trawl surveys on the Atlantic coast of Canada. Atlantic Zonal Monitoring Program Bull. 6: 25–42.

Dunn, P.K., and Smyth, G.K. 1996. Randomized quantile residuals. J Comput Graph Stat. 5: 236–44.

Duplisea, D.E., Bourdages, H., Brassard, C., Gauthier, J., Lambert, Y., Nitschke, P., and Valentin, A. 2016. Fitting a statistical catch at length model (NFT-SCALE) to Unit 1 + 2 redfish (*Sebastes mentella* and *Sebastes fasciatus*). DFO Can. Sci. Advis. Sec. Res. Doc. 2016/095. v + 32 p.

Gascon, D., Gagnon, P., Bernier, B., and Savard, L. 1991. Le relevé conjoint crevette/poisson de fond du nord du golfe du Saint-Laurent (divisions de l'OPANO 4RST). CSCPCA Document de travail 91/70 (unpublished working paper).

Gauthier, J., Marquis, M.-C., Bourdages, H., Ouellette-Plante, J. and Nozères, C. 2020. Gulf of St. Lawrence (4RST) Greenland Halibut Stock Status in 2018: Commercial Fishery and Research Survey Data. DFO Can. Sci. Advis. Sec. Res. Doc. 2020/016. v + 127 p.

Green, P.J., and Silverman, B.W. 1993. Nonparametric regression and generalized linear models. Chapman and Hall/CRC, 184 p.

Hastie, T., Tibshirani, R. and Friedman, J. 2009. The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media.

Hilborn, R., and Walters, C.J. 1992. Quantitative fisheries stock assessment: choice, dynamics and uncertainty. Chapman and Hall, New York

Hurlbut, T., and Clay, D. 1990. Protocols for research vessel cruises within the Gulf Region (demersal fish) (1970–1987). Can. Manuscr. Rep. Fish. Aquat. Sci. 2082.

Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H., and Bell, B.M. 2016. TMB: Automatic differentiation and Laplace approximation. J. Stat. Softw. 70: 1-21.

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.

Maunder, M.N., and Punt, A.E. 2013. A review of integrated analysis in fisheries stock assessment. Fish. Res. 142: 61–74

McAllister, M. and Duplisea, D.E. 2016. An updated production model fitting for redfish (*Sebastes fasciatus* and *Sebastes mentella*) in Units 1 and 2. DFO Can. Sci. Advis. Sec. Res. Doc. 2016/084. iv + 6 p.

Miller, T.J. 2013. A comparison of hierarchical models for relative catch efficiency based on paired-gear data for US Northwest Atlantic fish stocks. Can. J. Fish. Aquat. Sci. 70: 1306-1316.

Morin, R., Ricard, D., Benoît, H., and Surette, T. 2017. A review of the biology of Atlantic hagfish (*Myxine glutinosa*), its ecology, and its exploratory fishery in the southern Gulf of St. Lawrence (NAFO Div. 4T). DFO Can. Sci. Advis. Sec. Res. Doc. 2017/017. v + 39 p.

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

Ricard, D., and Swain, D.P. 2018. Assessment of Witch Flounder (*Glyptocephalus cynoglossus*) in the Gulf of St. Lawrence (NAFO Divisions 4RST), February 2017. DFO Can. Sci. Advis. Sec. Res. Doc. 2018/023. xi + 78 p.

Savenkoff, C., Castonguay, M., Chabot, D., Hammill, M.O., Bourdages, H., and Morissette, L. 2007a. Changes in the northern Gulf of St. Lawrence ecosystem estimated by inverse modelling: Evidence of a fishery-induced regime shift? Estuar. Coast. Shelf Sci. 73: 711-724.

Savenkoff, C., Swain, D.P., Hanson, J.M., Castonguay, M., Hammill, M.O., Bourdages, H., Morissette, L., and Chabot, D. 2007b. Effects of fishing and predation in a heavily exploited ecosystem: comparing periods before and after the collapse of groundfish in the southern Gulf of St. Lawrence (Canada). Ecol. Model., 204: 115-128.

Savoie, L. 2016. Indices of abundance to 2014 for six groundfish species based on the September research vessel and August sentinel vessel bottom-trawl surveys in the southern Gulf of St. Lawrence. DFO Can. Sci. Advis. Sec. Res. Doc. 2015/085. v + 52 p.

Swain, D.P., Poirier, G.A, and Morin, R. 1998. Relative fishing efficiency for witch flounder of vessels and gears used in the August and September bottom-trawl surveys in the Gulf of St. Lawrence. DFO Can. Sci. Advis. Sec. Res. Doc. 1998/03.

Thorson, J.T. and Minto, C. 2015. Mixed effects: a unifying framework for statistical modelling in fisheries biology. ICES J. Mar. Sci. 72:1245-1256.

Verbyla, A.P., Cullis, B.R., Kenward, M.G, and Welham, S.J. 1999. The analysis of designed experiments and longitudinal data by using smoothing splines. J. Roy. Stat. Soc. Ser. C 48: 269-311.

Wood, S.N. 2000. Modelling and smoothing parameter estimation with multiple quadratic penalties. J. R. Stat. Soc. Ser. B Stat. Methodol. 62: 413–428.

Wood, S.N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. J. R. Stat. Soc. Ser. B Stat. Methodol. 73: 3–36.

Wood, S.N. 2017. Generalized additive models: An introduction with R, 2nd ed. Chapman and Hall/CRC Press, 496 p.

Yin, Y. and Benoît, H.P. 2022. Length-specific relative catchabilities of redfish and Atlantic halibut by vessels and bottom trawls in multispecies research surveys in the Gulf of St. Lawrence based on paired-tow comparative fishing and spatio-temporal overlap. Can. Tech. Rep. Fish. Aquat. Sci. 3454: vi + 79 p.

6. TABLES

Table 1. Parameters for the vessels and summary of the protocols used in the RV surveys of the southern Gulf of St. Lawrence (sGSL) and northern Gulf of St. Lawrence (nGSL).

	E.E. Prince	*Lady Hammond*	*CCGS Alfred Needler*	*CCGS Teleost*
Regular survey operation	sGSL: 1971-1985	sGSL: 1985-1991 nGSL: 1984-1990	sGSL: 1992-2005 nGSL: 1990-2005	sGSL: 2004-present nGSL: 2004-present
Vessel type	Stern trawler	Stern trawler	Stern trawler	Stern trawler
Tonnage	406	897	959	2,405
Length (m)	40	58	50	63
Operating hours	Daylight only (7:00-19:00)	24-hr	24-hr	24-hr
Standard tow speed (knots)	3.5	3.5	sGSL: 3.5 nGSL (1990-1993): 2.5 nGSL (1994-2005): 3.0	sGSL: 3.5 nGSL: 3.0
Standard tow duration (min)	30	30	sGSL: 30 nGSL (1990-1992): 20 nGSL (1993-2005): 24	sGSL: 30 nGSL: 15
Standard tow distance (nm)	1.75	1.75	sGSL: 1.75 nGSL (1990-1992): 0.83 nGSL (1993): 1.00 nGSL (1994-2005): 1.20	sGSL: 1.75 nGSL: 0.75

Table 2. Parameters for the trawls used in the RV surveys of the southern Gulf of St. Lawrence (sGSL) and northern Gulf of St. Lawrence (nGSL).

	Yankee 36	Western IIA	URI 81/114	Campelen
Years in operation	sGSL: 1971-1984	sGSL: 1985-present nGSL: 1984-1990	nGSL: 1990-2005	nGSL: 2004-present
Footgear	7 inch (outer sections) and 14 inch (inner sections) rubber disc spacers + 17 lb. iron spacers	21 inch (outer) and 18 inch (inner) rubber bobbins and 6.75 inch diameter 7 inch long rubber spacers	information not available	Rockhopper
Footrope length (m)	24.4	32.3	34.8	35.6
Headline length (m)	18.3	22.9	24.7	29.5
Headline height (m)	2.7	4.6	5.5	-
Wingspread (m)	10.7	12.5	14-15	16-17
Door type	Steel bound wood	Portuguese (all steel)	Morgère	Polyvalent
Lengthening piece liner (mm)	31.75	31.75	44.0	44.0
Codend liner (mm)	6.35	19.0	19.0	12.7

Table 3. A set of binomial models with various assumptions on the length effect and station effect in the relative catch efficiency. A smoothing length effect can be considered and the station effect can be added to the intercept, without interaction with the length effect, or added to both the intercept and smoother to allow for interaction between the two effects.

Model	$\boldsymbol{log(\rho)}$	Length Effect	Station Effect
$BI0$	β_0	constant	not considered
$BI1$	$\beta_0 + \delta_{0,i}$	constant	intercept
$BI2$	$\mathbf{X}_f^T\boldsymbol{\beta}_f + \mathbf{X}_r^T\mathbf{b}$	smoothing	not considered
$BI3$	$\mathbf{X}_f^T\boldsymbol{\beta}_f + \mathbf{X}_r^T\mathbf{b} + \delta_{0,i}$	smoothing	intercept
$BI4$	$\mathbf{X}_f^T(\boldsymbol{\beta}_f + \boldsymbol{\delta}_i) + \mathbf{X}_r^T(\mathbf{b} + \boldsymbol{\epsilon}_i)$	smoothing	intercept, smoother

Table 4. A set of beta-binomial models with various assumptions on the length effect and station effect in the relative catch efficiency, and the length effect on the variance parameter. A smoothing length effect can be considered in both the conversion factor and the variance parameter. A possible station effect can be added to the intercept, without interaction with the length effect, or added to both the intercept and the smoother to allow for interaction between the two effects.

Model	$\boldsymbol{log(\rho)}$	$\boldsymbol{log(\phi)}$	Length Effects	Station Effect
BB_0	β_0	γ_0	constant/constant	not considered
BB_1	$\beta_0 + \delta_{0,i}$	γ_0	constant/constant	intercept
BB_2	$\mathrm{X}_f^T\beta_f + \mathrm{X}_r^T\mathrm{b}$	γ_0	smoothing/constant	not considered
BB_3	$\mathrm{X}_f^T\beta_f + \mathrm{X}_r^T\mathrm{b}$	$\mathrm{X}_f^T\gamma + \mathrm{X}_r^T\mathrm{g}$	smoothing/smoothing	not considered
BB_4	$\mathrm{X}_f^T\beta_f + \mathrm{X}_r^T\mathrm{b} + \delta_{0,i}$	γ_0	smoothing/constant	intercept
BB_5	$\mathrm{X}_f^T\beta_f + \mathrm{X}_r^T\mathrm{b} + \delta_{0,i}$	$\mathrm{X}_f^T\gamma + \mathrm{X}_r^T\mathrm{g}$	smoothing/smoothing	intercept
BB_6	$\mathrm{X}_f^T(\beta_f + \delta_i) + \mathrm{X}_r^T(\mathrm{b} + \epsilon_i)$	γ_0	smoothing/constant	intercept, smoother
BB_7	$\mathrm{X}_f^T(\beta_f + \delta_i) + \mathrm{X}_r^T(\mathrm{b} + \epsilon_i)$	$\mathrm{X}_f^T\gamma + \mathrm{X}_r^T\mathrm{g}$	smoothing/smoothing	intercept, smoother

Table 5. Difference in AIC from the model giving lowest AIC for each of the 13 candidate models and each comparative fishing analysis. The best model (indicated in bold) was selected by the lowest AIC. However, in sGSL 1985, the best model was selected by the lowest AIC among models without length effect (BI0, BI1 and BB1). A dash indicates a model that was either not attempted due to lack of data or for which convergence could not be achieved.

Model	nGSL1990	nGSL.2004.2005	sGSL.1985	sGSL.1992	sGSL.2004.2005
BI0	162	1068	105	228	278
BI1	179	192	114	**0**	209
BI2	**0**	987	93	97	140
BI3	17	135	97	-	77
BI4	-	-	-	-	-
BB0	158	639	**105**	230	248
BB1	-	160	-	5	211
BB2	2	587	94	120	114
BB3	-	463	0	-	**0**
BB4	-	107	-	-	79
BB5	-	**0**	-	-	-
BB6	-	-	-	-	-
BB7	-	-	-	-	-

7. FIGURES

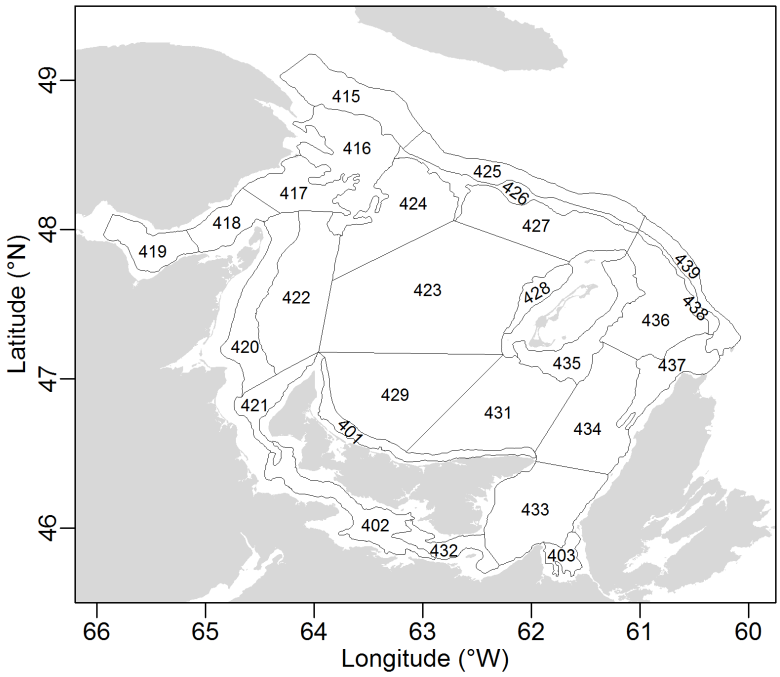

Figure 1. Stratification scheme for the southern Gulf of St. Lawrence multi-species bottom-trawl survey.

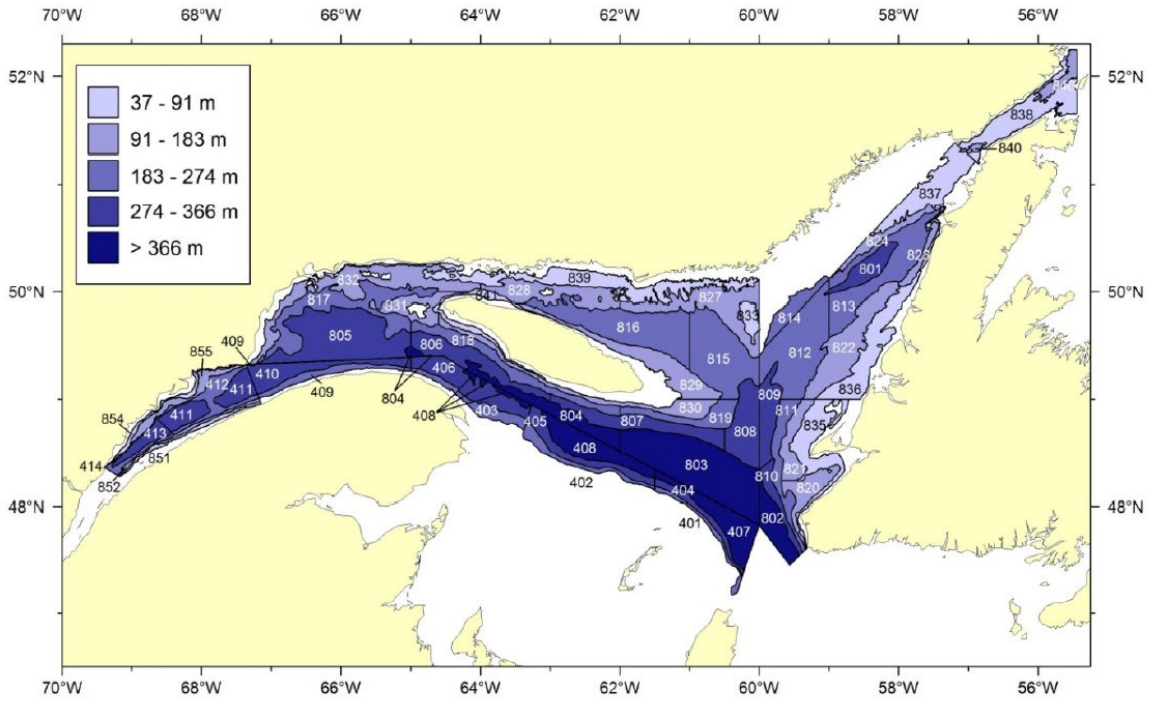

Figure 2. Stratification scheme for the northern Gulf of St. Lawrence multi-species bottom-trawl survey. Strata 401-408, 801-824 and 827-832 constitute a core group of strata included annually in the sampling design since at least 1985. Additional strata, located in NAFO area 3Ps (southwest Newfoundland) and sampled only in 1987 and 1993-2003 are not shown.

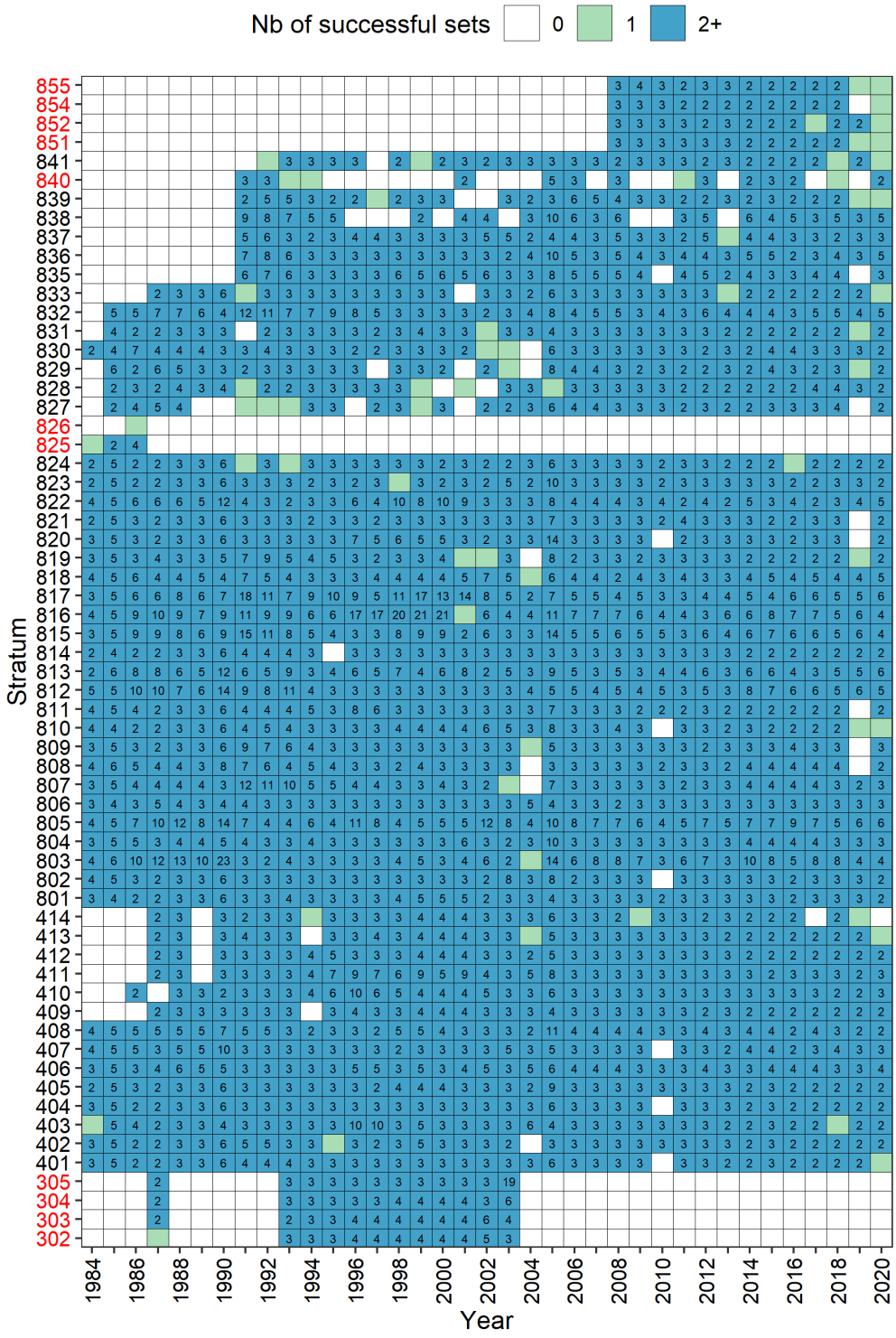

Figure 3. Summary of the number of survey sets made in each stratum and year in the northern Gulf of St. Lawrence survey. Strata indicated in red are not used in the estimation of survey abundance series for any taxa because of inconsistent sampling over the years. (Figure courtesy of Jordan Ouellette-Plante, DFO Québec region).

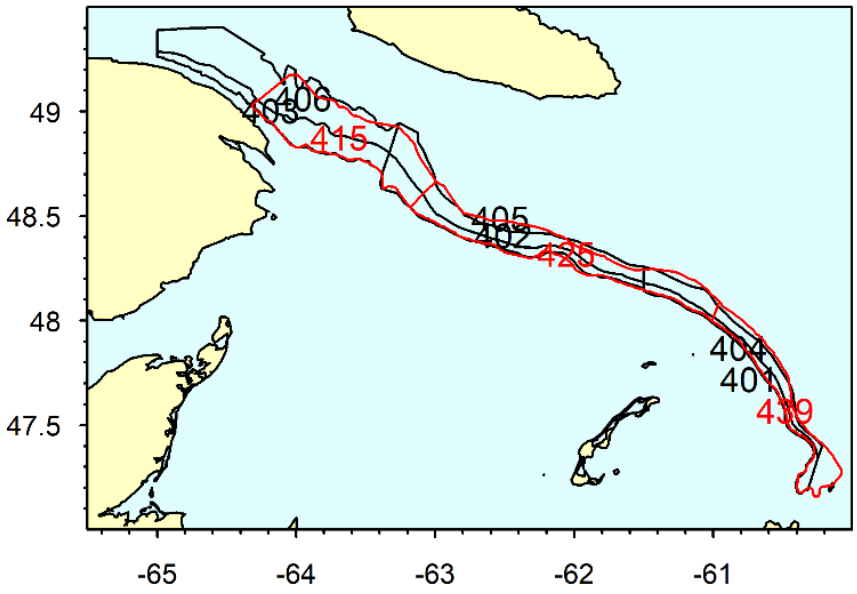

Figure 4. A close-up map of the strata from the southern Gulf of St. Lawrence (in red; strata 415, 425, 439) and northern Gulf of St. Lawrence (in black; strata 401-406) surveys in the area along the southern slope of the Laurentian channel where the two surveys overlap.

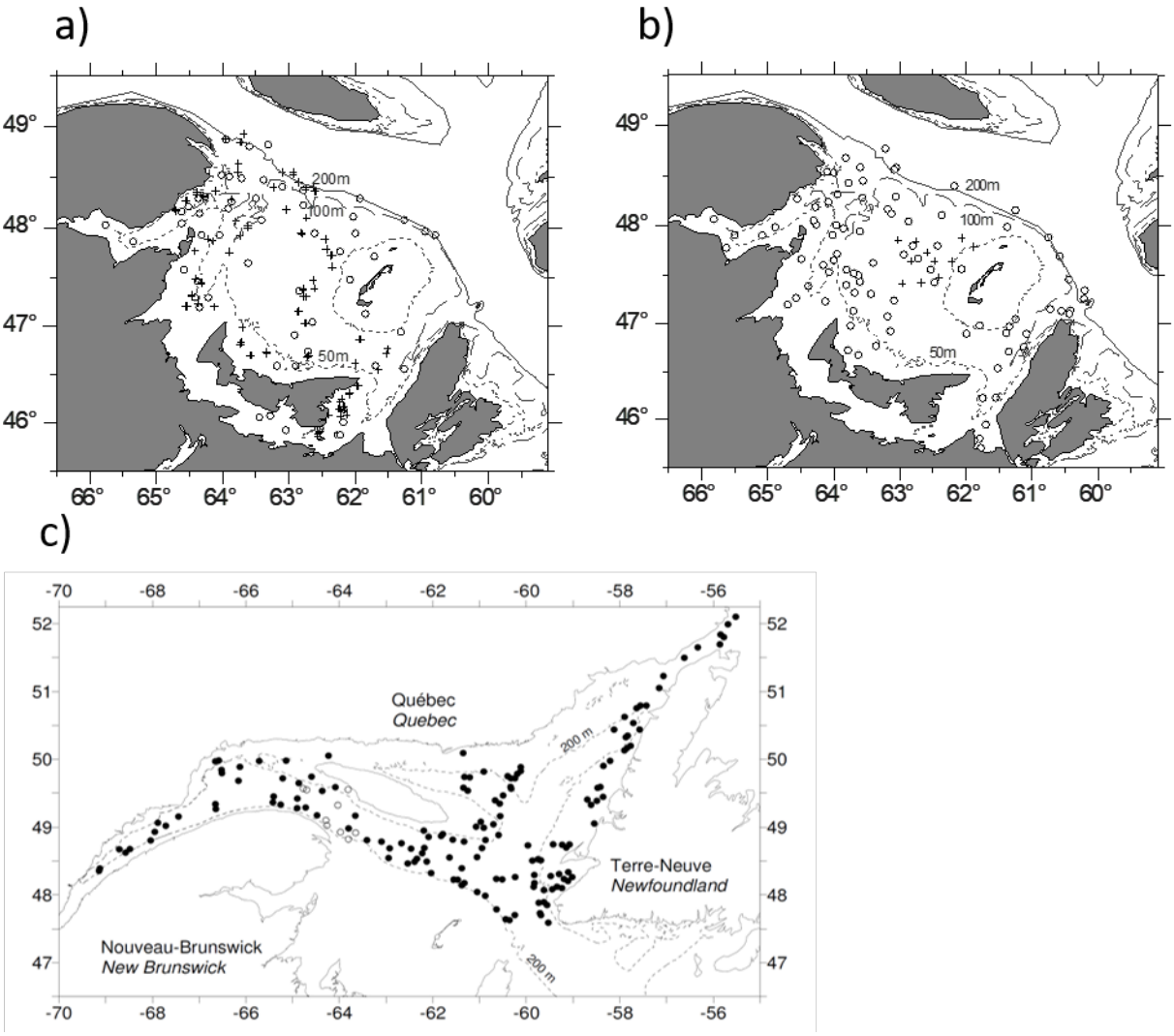

Figure 5. Location of comparative fishing sets in four comparative fishing experiments in the Gulf of St. Lawrence: a) 1985 (o) and 1992 (+) comparative fishing experiments in the southern Gulf survey (from Benoît and Swain 2003), b) 2004 (+) and 2005 (o) comparative fishing experiments in the southern Gulf survey (from Benoît 2006), and c) 2004 (o) and 2005 (•) comparative fishing experiments in the northern Gulf survey (from Bourdages et al. 2007).

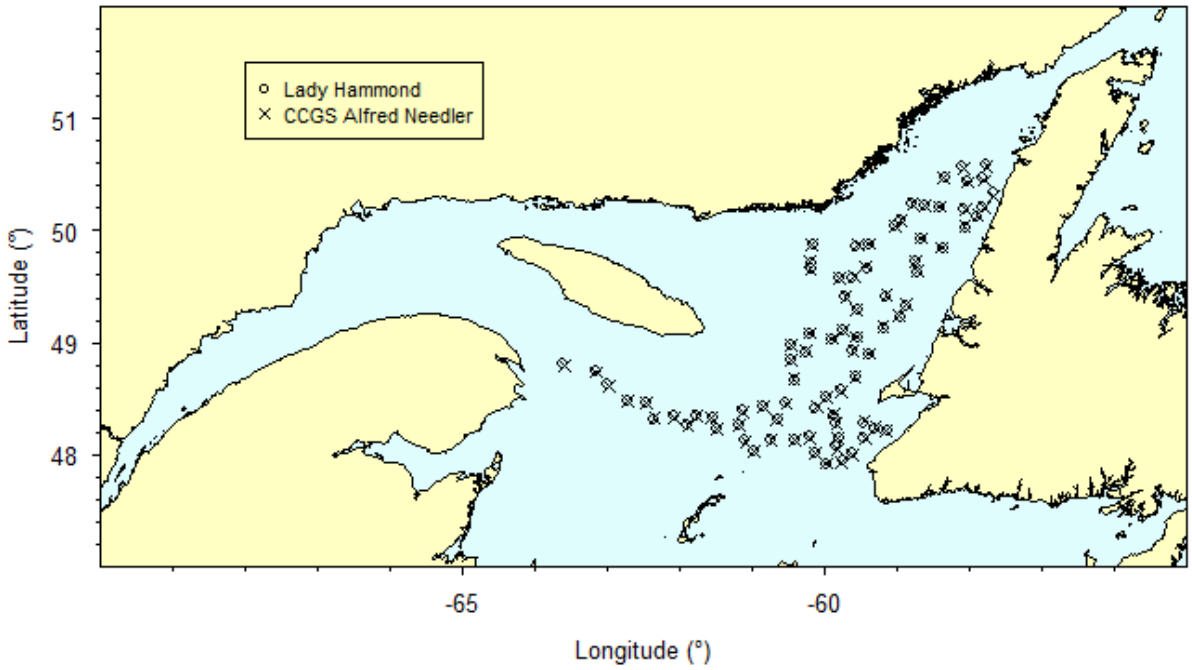

Figure 6. Location of successful comparative fishing tows during the August 1990 northern Gulf of St. Lawrence bottom-trawl survey.

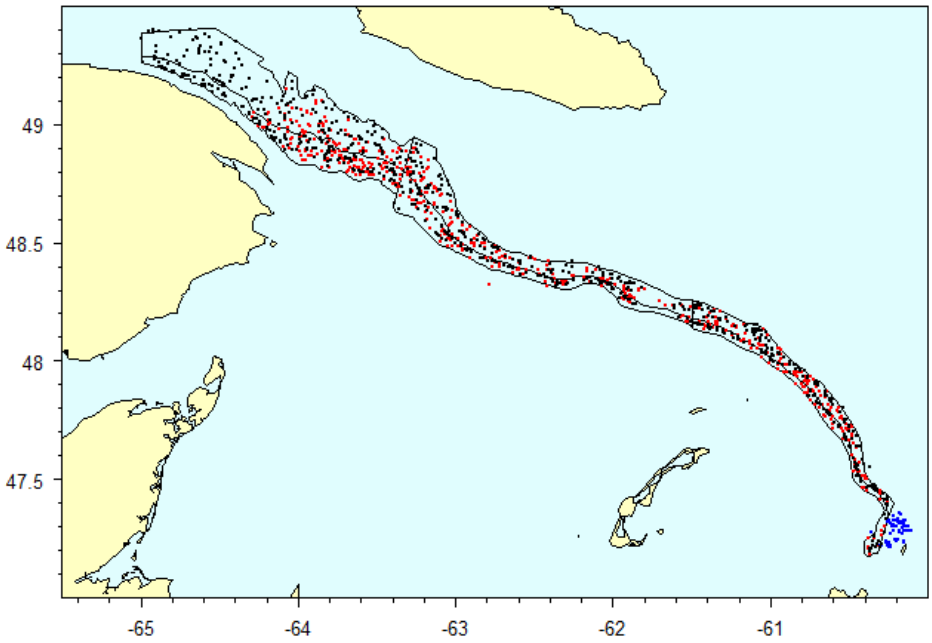

Figure 7. Location of successful fishing sets by the northern Gulf survey (black points) and the southern Gulf survey (red points) in nGSL strata 401-406. Also shown are locations of survey sets by the southern Gulf survey in stratum 439, which do not fall in nGSL strata 401-406 and which were not retained in the analyses (blue points).

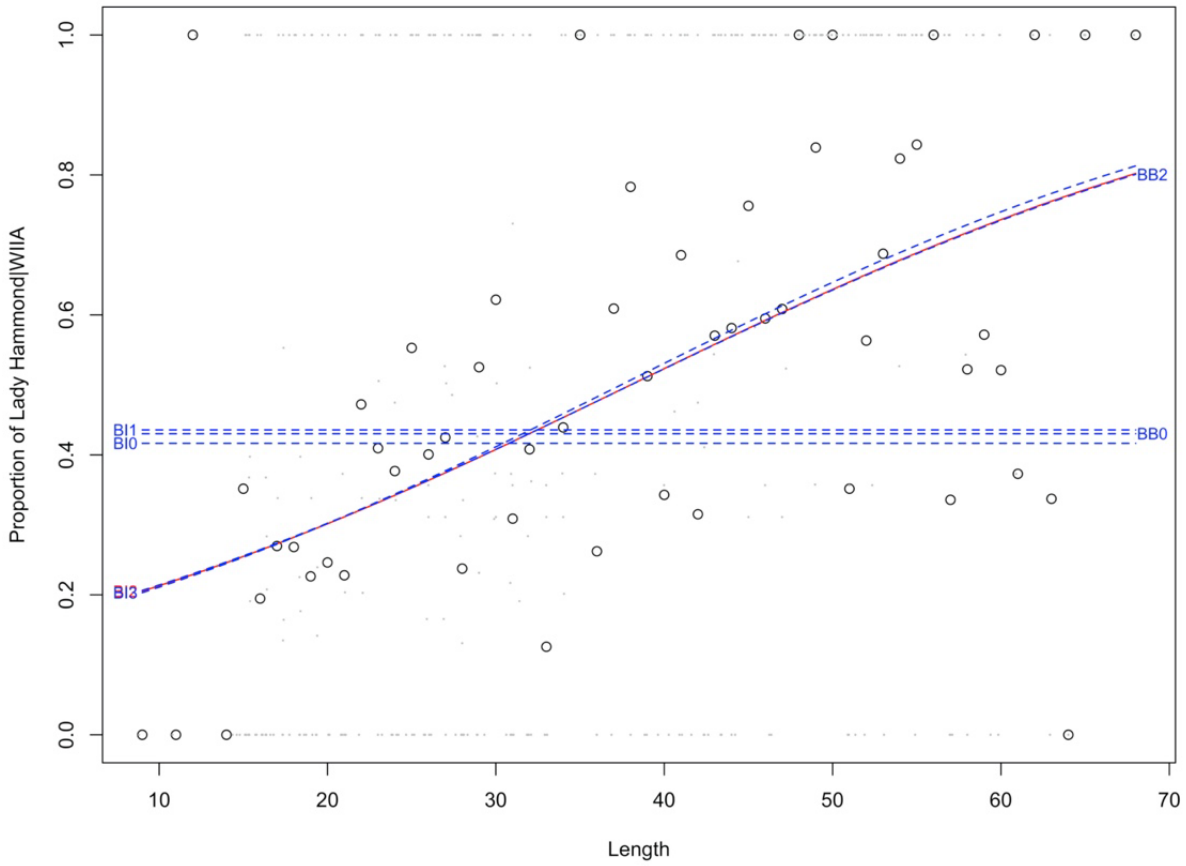

Figure 8. Comparative fishing analysis of the Lady Hammond-WIIA and Alfred Needler-URI catches in the nGSL during 1990: Estimated proportion of catch over length by the Lady Hammond-WIIA from the candidate binomial and beta-binomial models (red solid line for the selected best model and blue dashed lines for other converged models), compared to the sample proportion of catch by length (gray dots for each paired tow within each station and black circles for the average across stations).

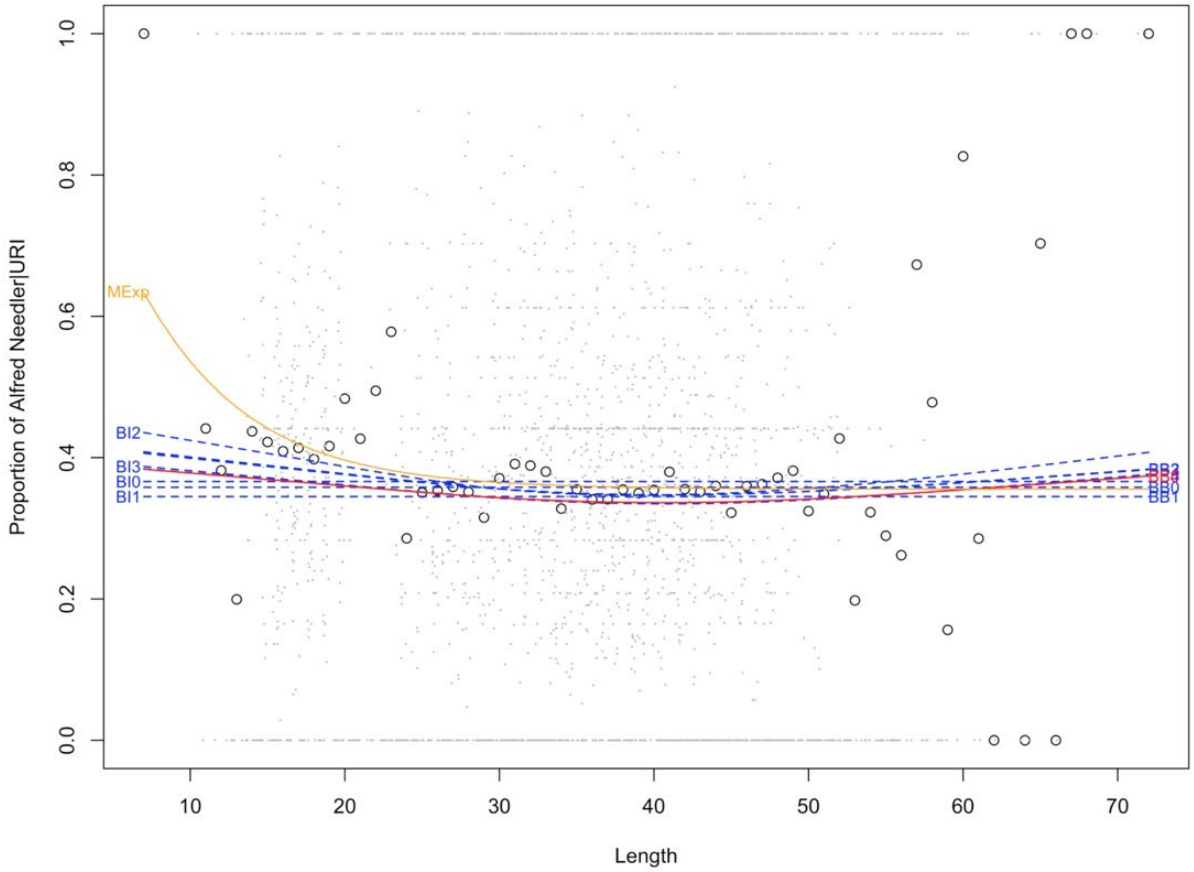

Figure 9. Comparative fishing analysis of the Alfred Needler-URI and Teleost-Campelen catches in the nGSL during 2004-2005: Estimated proportion of catch over length by the Alfred Needler-URI from the candidate binomial and beta-binomial models (red solid line for the selected best model and blue dashed lines for other converged models) and from the exponential model by Bourdages et al. (2007) (yellow solid line), compared to the sample proportion of catch by length (gray dots for each paired tow within each station and black circles for the average across stations).

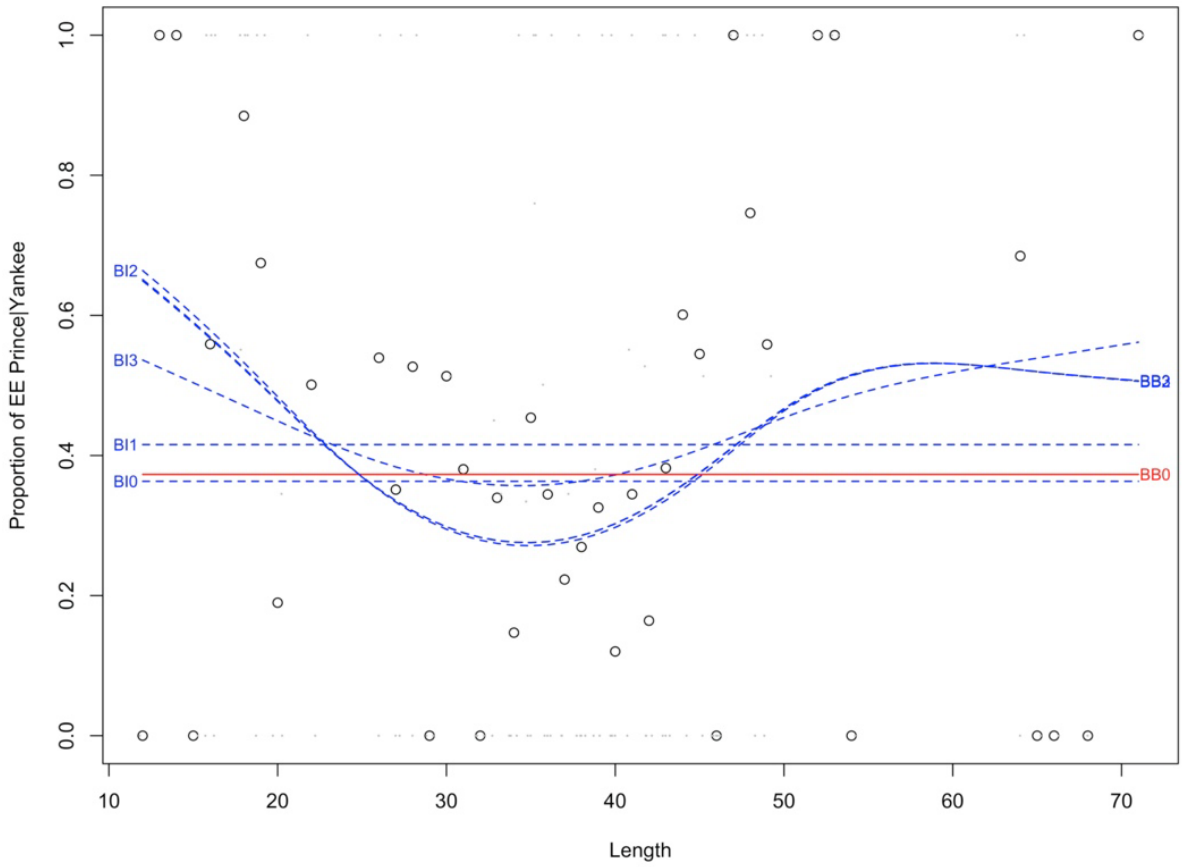

Figure 10. Comparative fishing analysis of the EE Prince-Yankee and Lady Hammond-WIIA catches in the sGSL during 1985: Estimated proportion of catch over length by the EE Prince-Yankee from the candidate binomial and beta-binomial models (red solid line for the selected best model and blue dashed lines for other converged models), compared to the sample proportion of catch by length (gray dots for each paired tow within each station and black circles for the average across stations). The best model was selected by the lowest AIC among models without length effect.

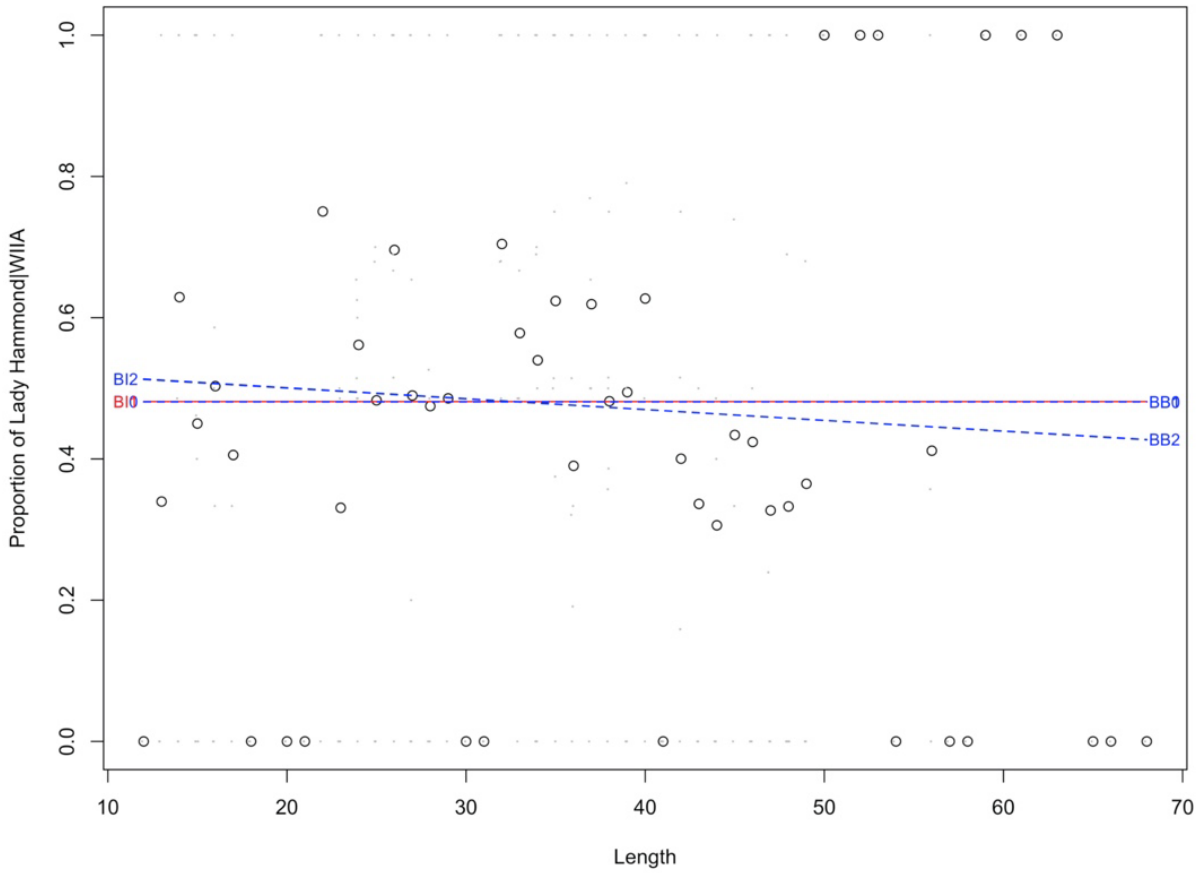

Figure 11. Comparative fishing analysis of the Lady Hammond-WIIA and Alfred Needler-WIIA catches in the sGSL during 1992: Estimated proportion of catch over length by the Lady Hammond-WIIA from the candidate models (red solid line for the selected best model and blue dashed lines for other converged models), compared to the sample proportion of catch by length (gray dots for each paired tow within each station and black circles for the average across stations).

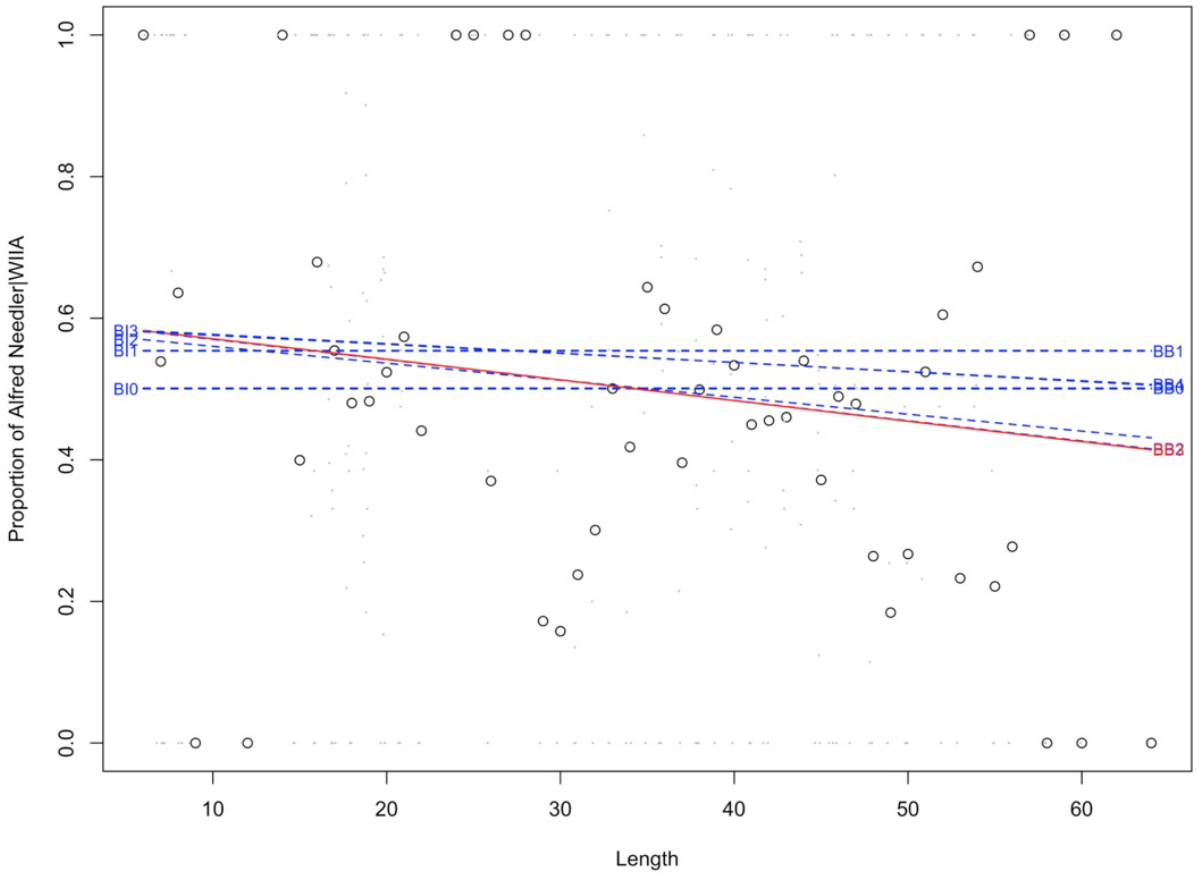

Figure 12. Comparative fishing analysis of the Alfred Needler-WIIA and Teleost-WIIA catches in the sGSL during 2004-2005: Estimated proportion of catch over length by the Alfred Needler-WIIA from the candidate models (red solid line for the selected best model and blue dashed lines for other converged models), compared to the sample proportion of catch by length (gray dots for each paired tow within each station and black circles for the average across stations).

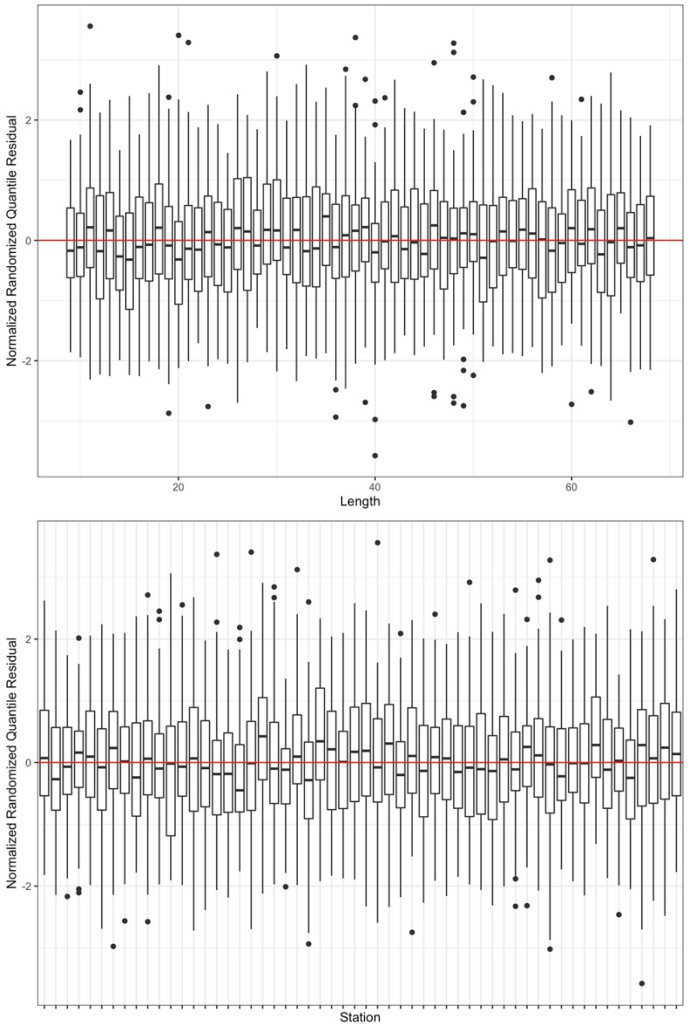

Figure 13. Comparative fishing analysis of the Lady Hammond-WIIA and Alfred Needler-URI catches in the nGSL during 1990: normalized randomized quantile residuals for each length bin (top panel) and for each station (bottom panel) did not indicate significant deviations.

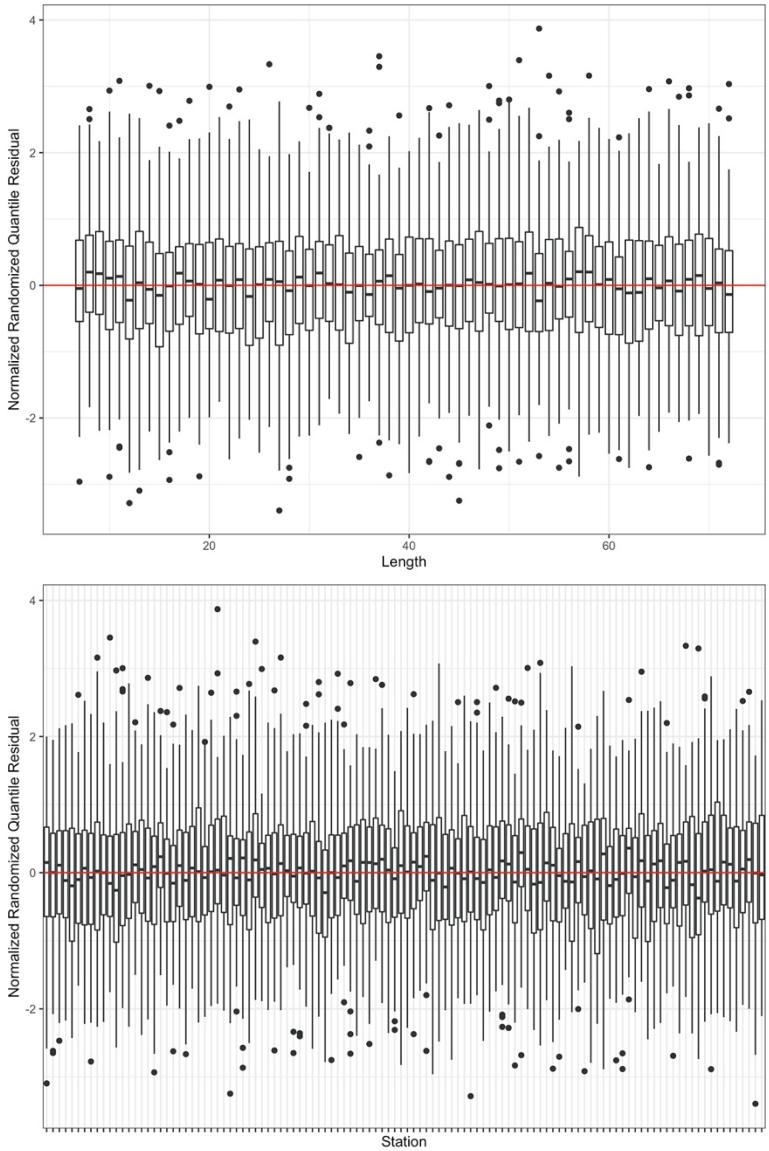

Figure 14. Comparative fishing analysis of the Alfred Needler-URI and Teleost-Campelen catches in the nGSL during 2004-2005: normalized randomized quantile residuals for each length bin (top panel) and for each station (bottom panel) did not indicate significant deviations.

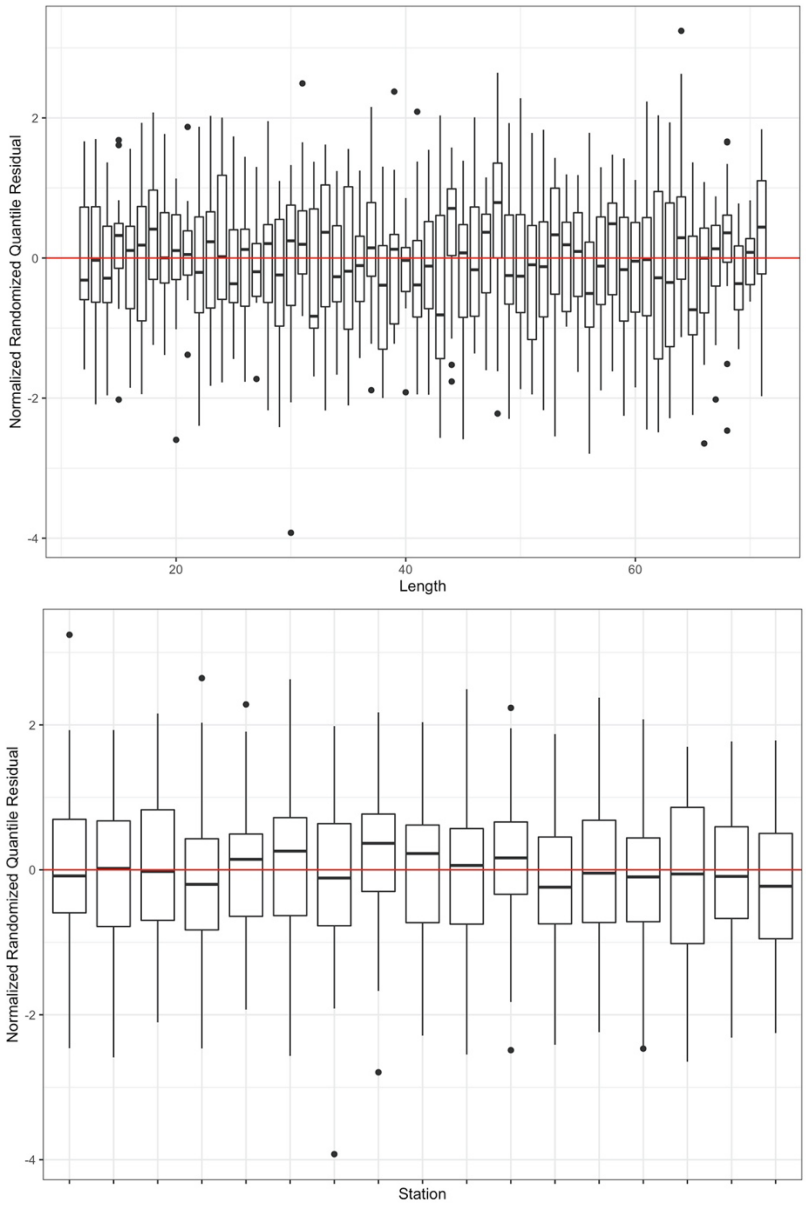

Figure 15. Comparative fishing analysis of the EE Prince-Yankee and Lady Hammond-WIIA catches in the sGSL during 1985: normalized randomized quantile residuals for each length bin (top panel) and for each station (bottom panel. The bias is higher for some lengths due to small sample size (number of effective observations).

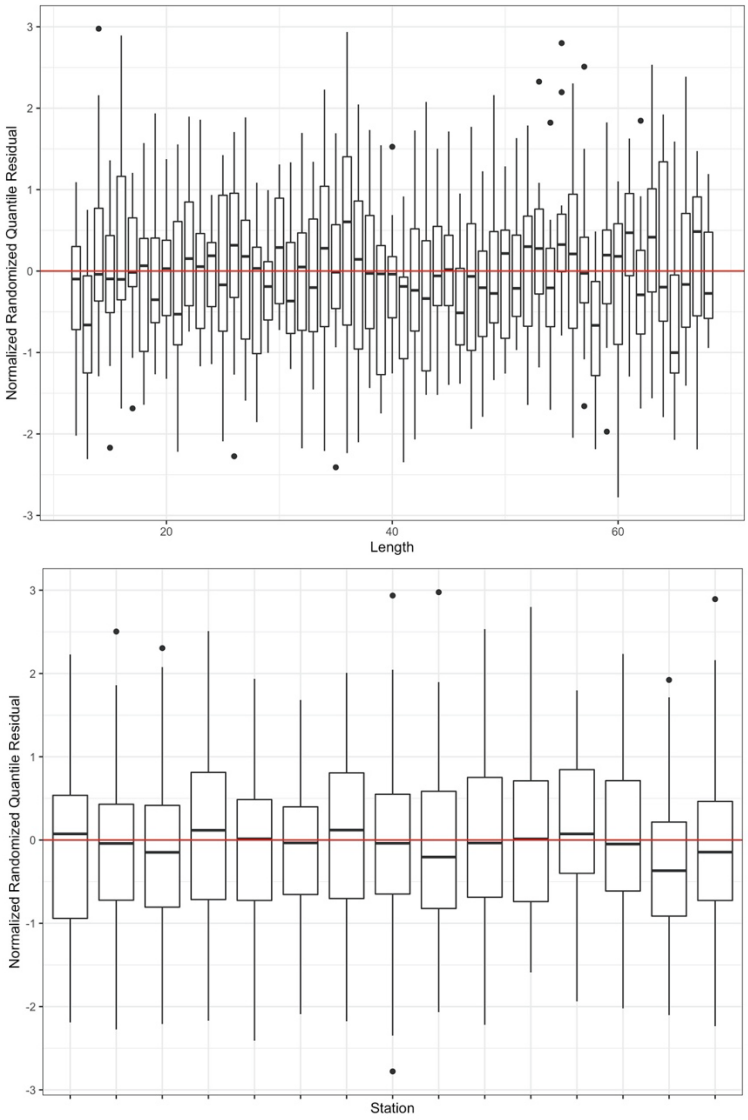

Figure 16. Comparative fishing analysis of the Lady Hammond-WIIA and Alfred Needler-WIIA catches in the sGSL during 1992: normalized randomized quantile residuals for each length bin (top panel) and for each station (bottom panel). The bias is higher for some lengths due to small sample size (number of effective observations).

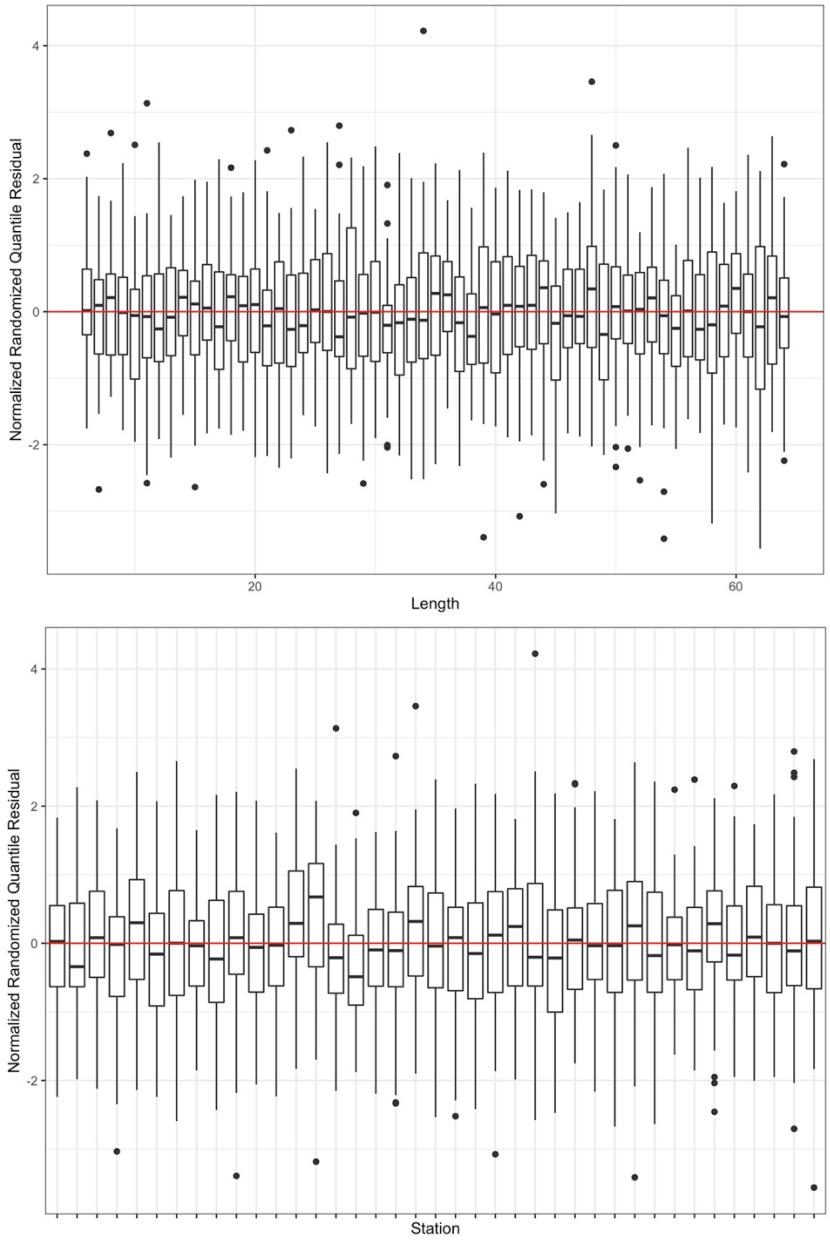

Figure 17. Comparative fishing analysis of the Alfred Needler-WIIA and Teleost-WIIA catches in the sGSL during 2004-2005: normalized randomized quantile residuals for each length bin (top panel) and for each station (bottom panel) did not indicate significant deviations.

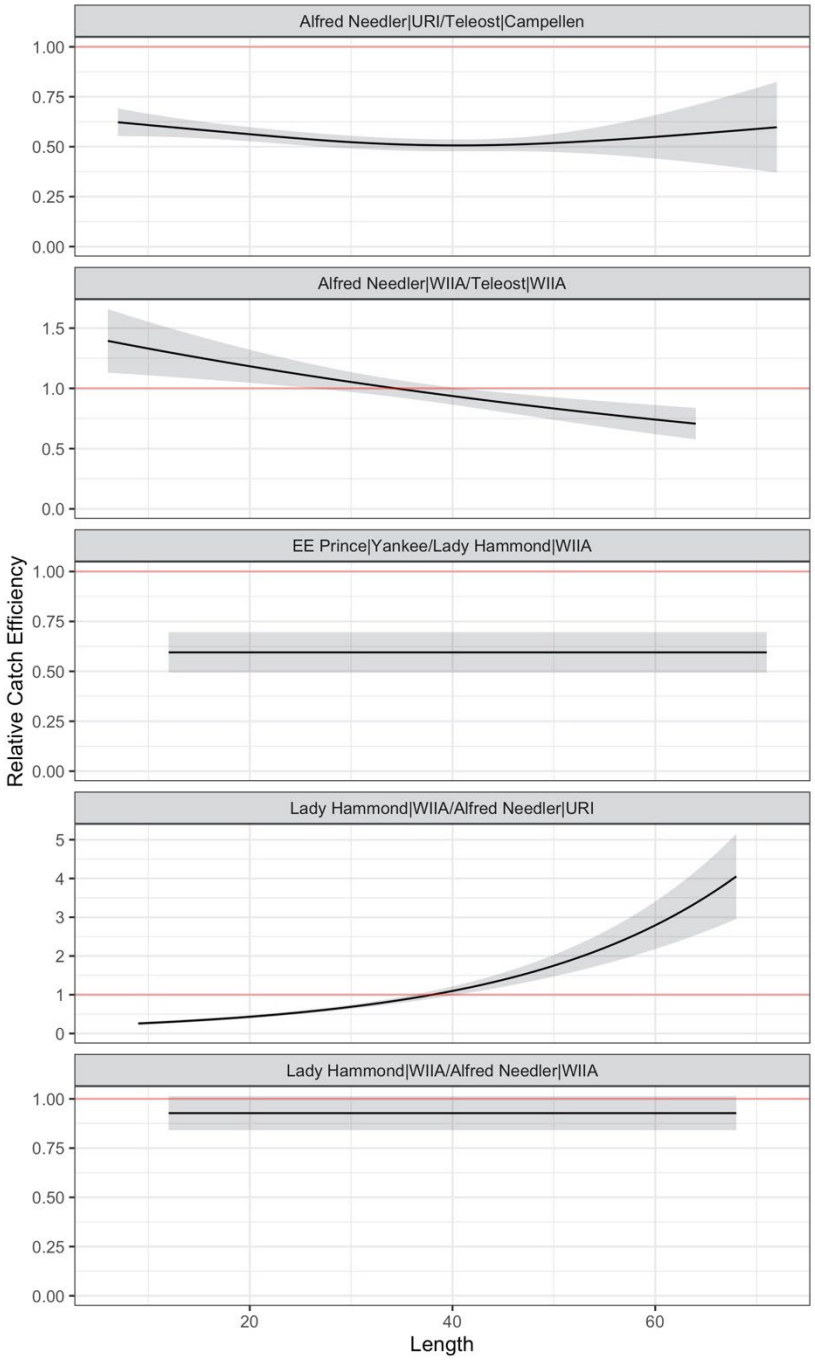

Figure 18. Estimated relative catch efficiency from each comparative fishing experiment. Shaded areas indicate plus/minus one standard error. The horizontal red line indicates equal efficiency.

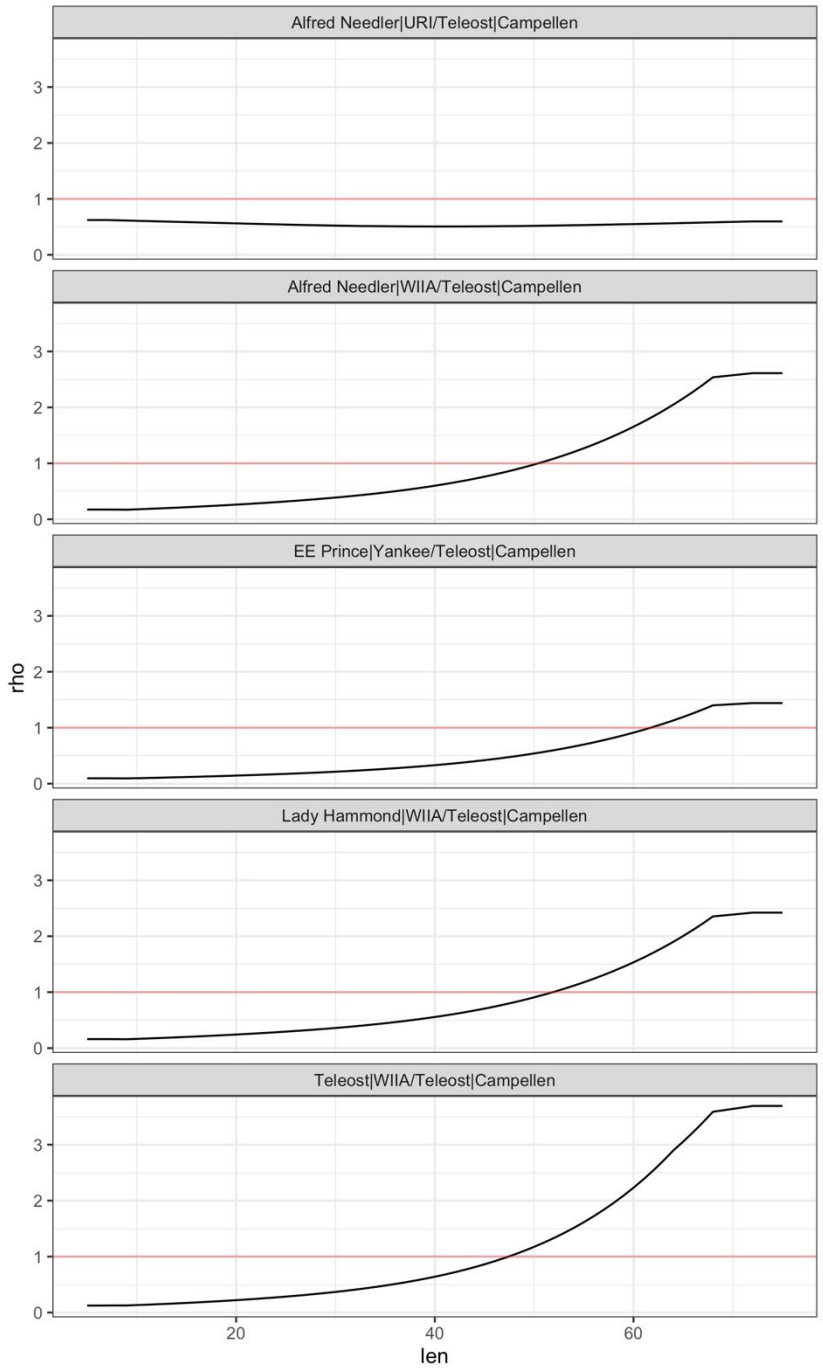

Figure 19. Estimated catch efficiency relative to Teleost-Campelen for each vessel-gear based on sequential multiplication of the estimated relative catch efficiencies from the five comparative fishing analyses. The horizontal red line indicates equal efficiency.

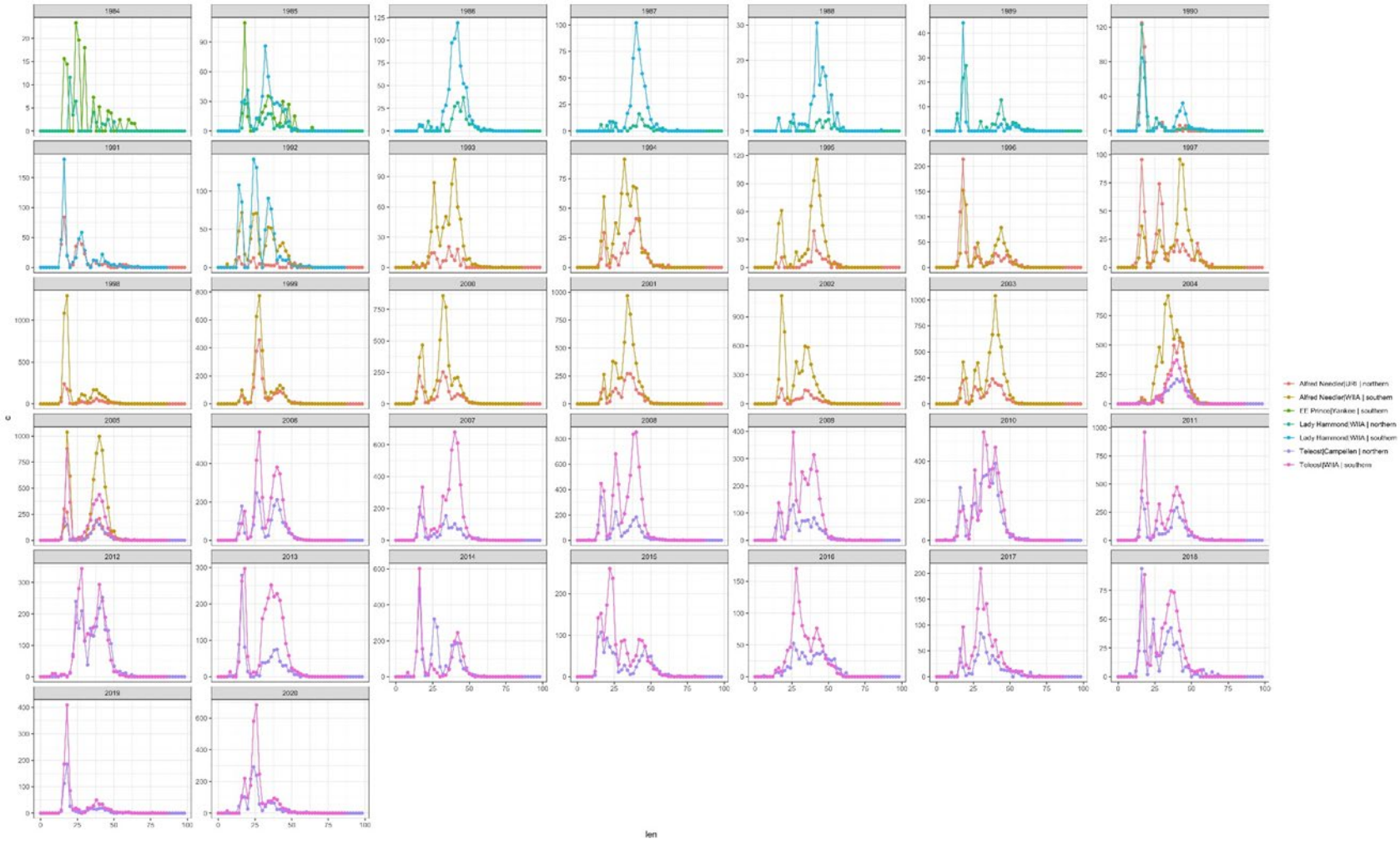

Figure 20. Survey-aggregated annual mean adjusted catches at length by each vessel, in each survey (distinguished by colour), for the area of survey overlap in the Laurentian channel.

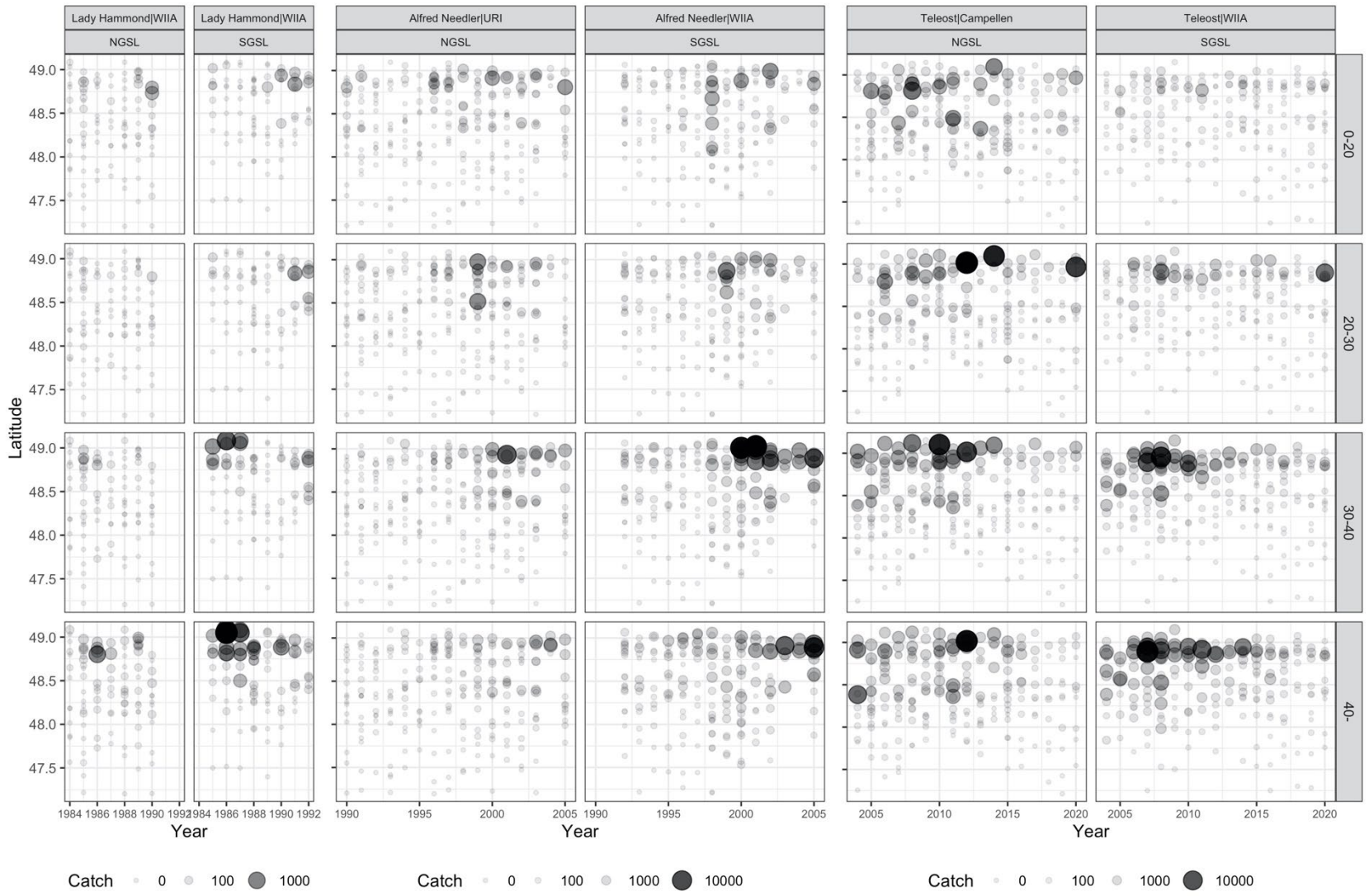

Figure 21. Standardized (adjusted) catches in individual nGSL and sGSL survey tows as a function of latitude and year, and grouped by length class and vessel-gear for the area of survey overlap in the Laurentian channel. Circle size and shading indicate magnitude of total catch; note the different scale used for different vessel-gear groups over the years, according to the legend at the bottom of the plots.

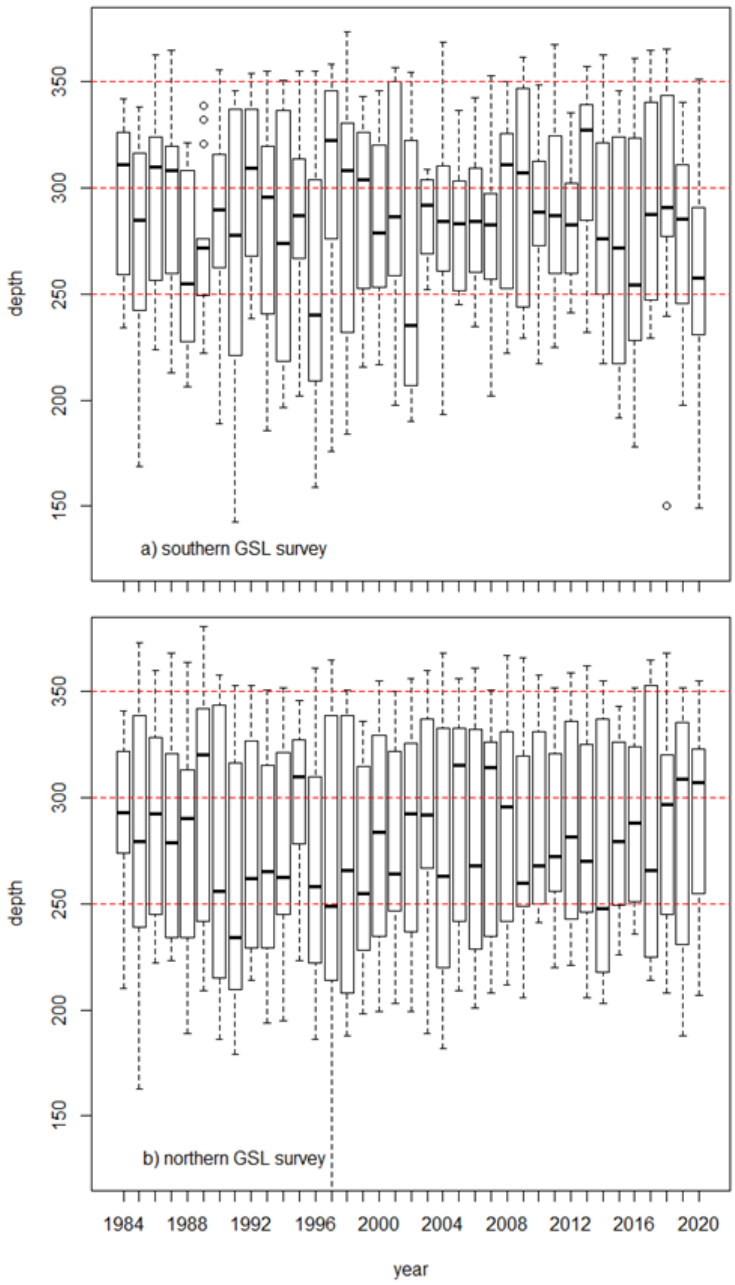

Figure 22. Boxplots of annual depths sampled by the a) southern and b) northern Gulf surveys in the area of overlap. Reference lines are drawn at 250, 300 and 350 m to facilitate comparisons.

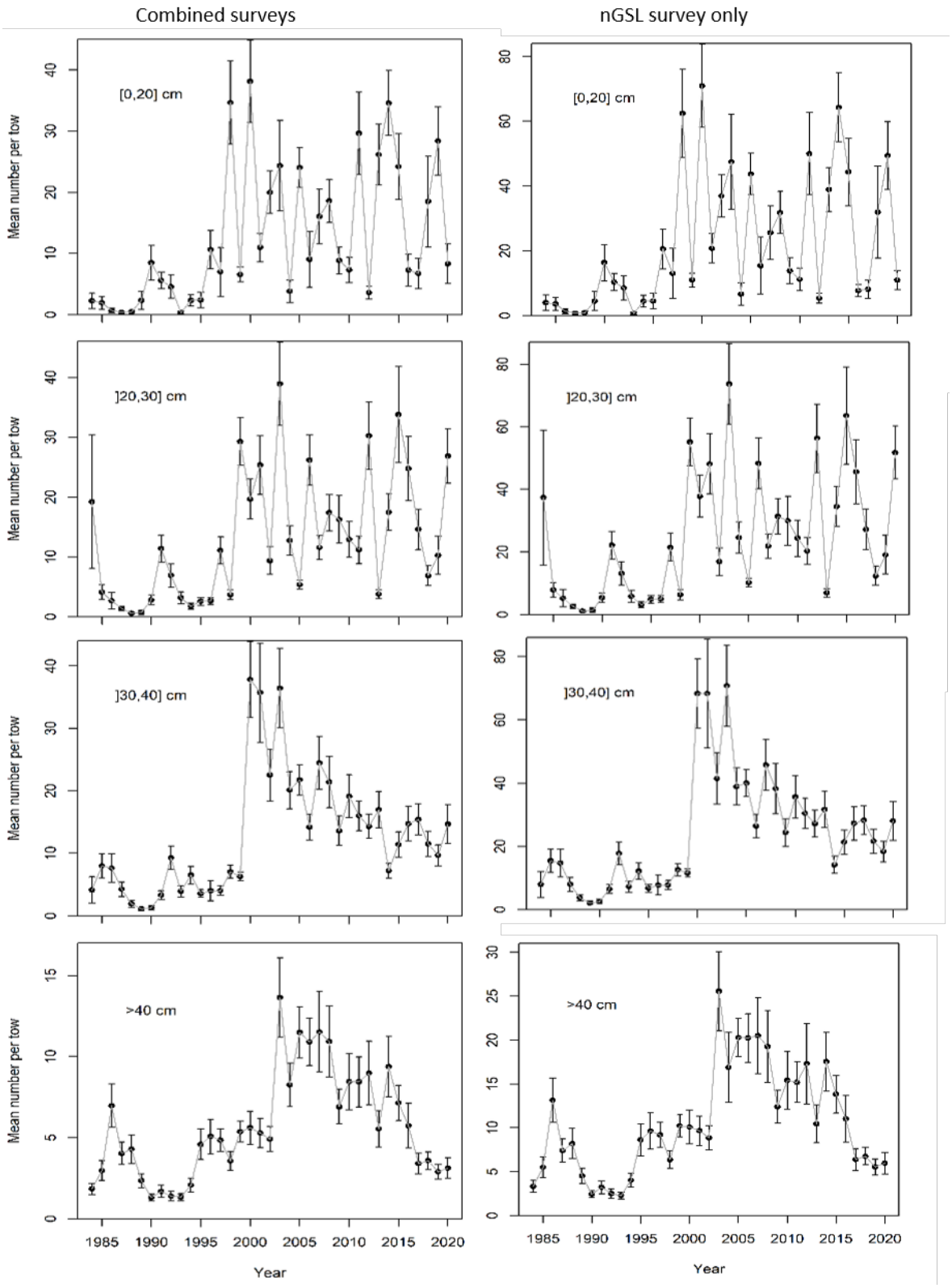

Figure 23. Abundance indices (mean number per tow) for Greenland halibut, by size class (rows), in combined nGSL and sGSL surveys (left column) and the nGSL surveys only (right column) for 1984-2020.

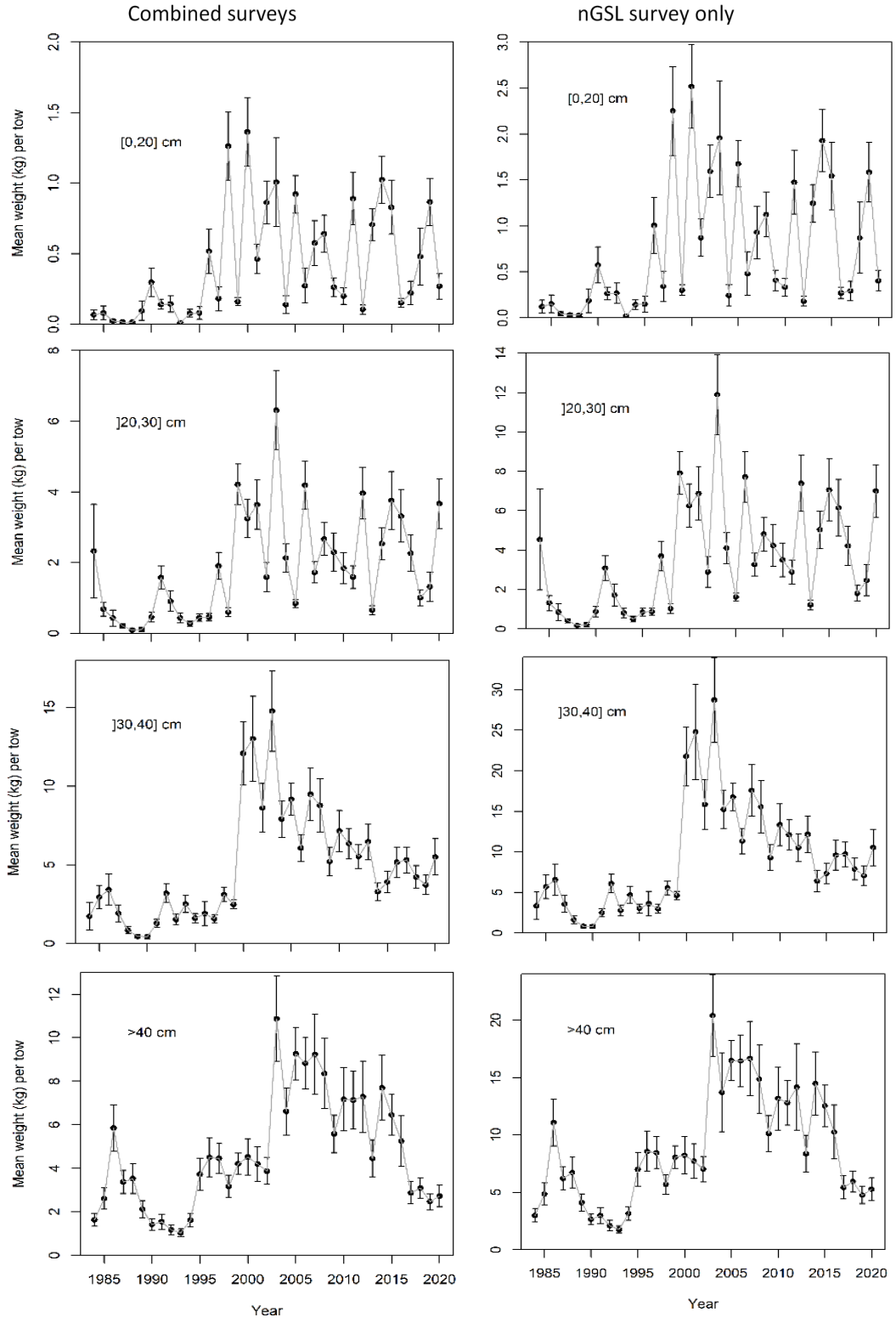

Figure 24. Biomass indices (mean kg per tow) for Greenland halibut, by size class (rows), in combined nGSL and sGSL surveys (left column) and the nGSL surveys only (right column) for 1984-2020.

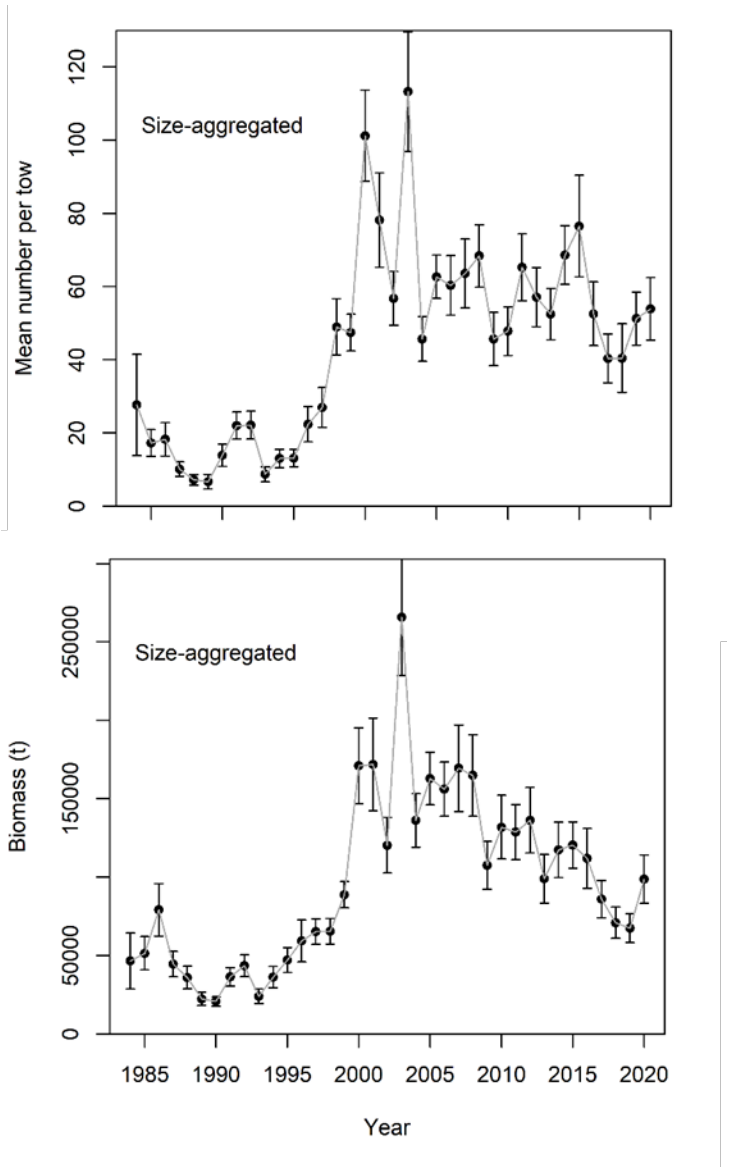

Figure 25. Size-aggregated abundance index (mean number per tow; top panel) and trawlable biomass (tonnes; bottom panel) for Greenland halibut in the combined nGSL and sGSL surveys, 1984-2020.

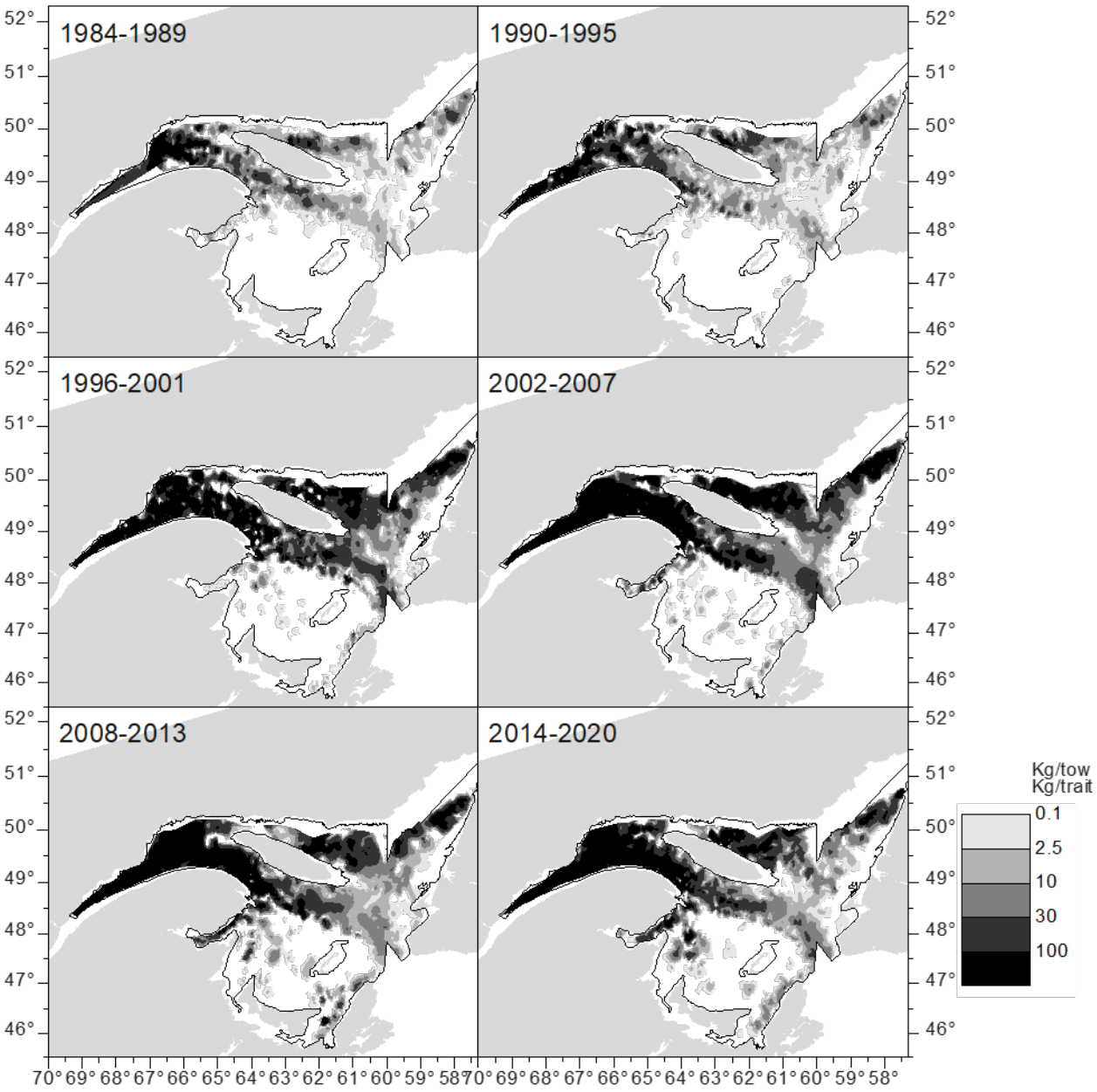

Figure 26. Distribution of Greenland halibut biomass (kg per tow; size-aggregated) in 6 or 7 year blocks in the joint nGSL and sGSL survey, 1984-2020. Interpolation is based on Delaunay triangles. To avoid the inappropriate formation of Delaunay triangles between distant points and points topologically separated by barriers, a blanking distance of 0.7 degrees was used as the distance limit between data points at which Delaunay triangles were removed. If one or more sides of a Delaunay triangle had a length that exceeded this value the triangle was not contoured.

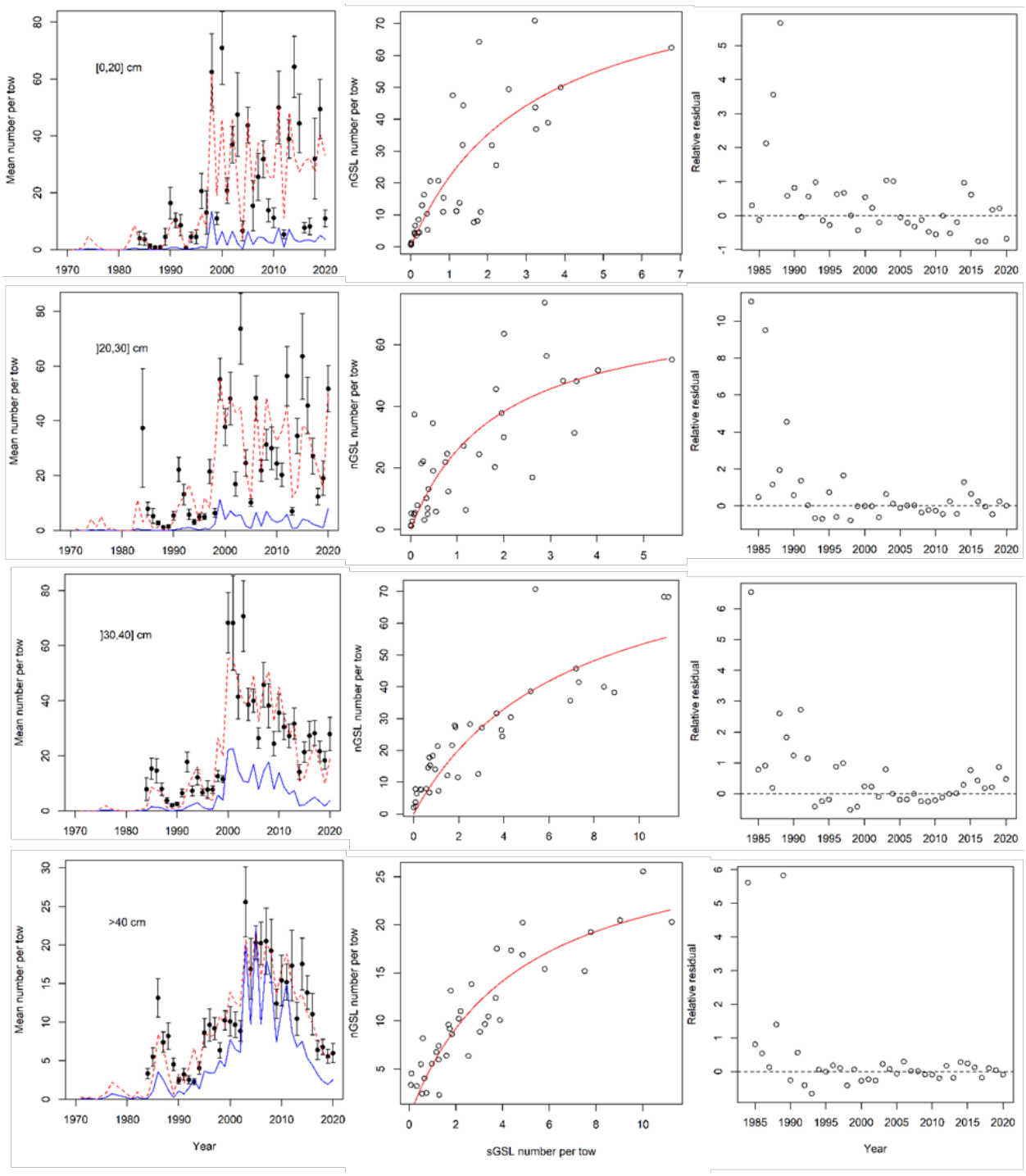

Figure 27. Length-group specific abundance indices (mean number per tow; left column) for the whole GSL index (black dots and 95% confidence interval), the sGSL index (blue line) and the whole GSL index predicted from the sGSL index using the density-dependent relationship (red dotted line). The panels in the middle column show the whole GSL indices (y-axis) as a function of the sGSL indices (x-axis), as well as the modelled density-dependent relationship between them (red line). The panels in the right column show the relative residuals from the density-dependent relationship, (observed-predicted)/predicted, as a function of year. Note that in the left panels the sGSL indices were arbitrarily multiplied by two to make their magnitude closer to those of the whole GSL indices for plotting purposes.

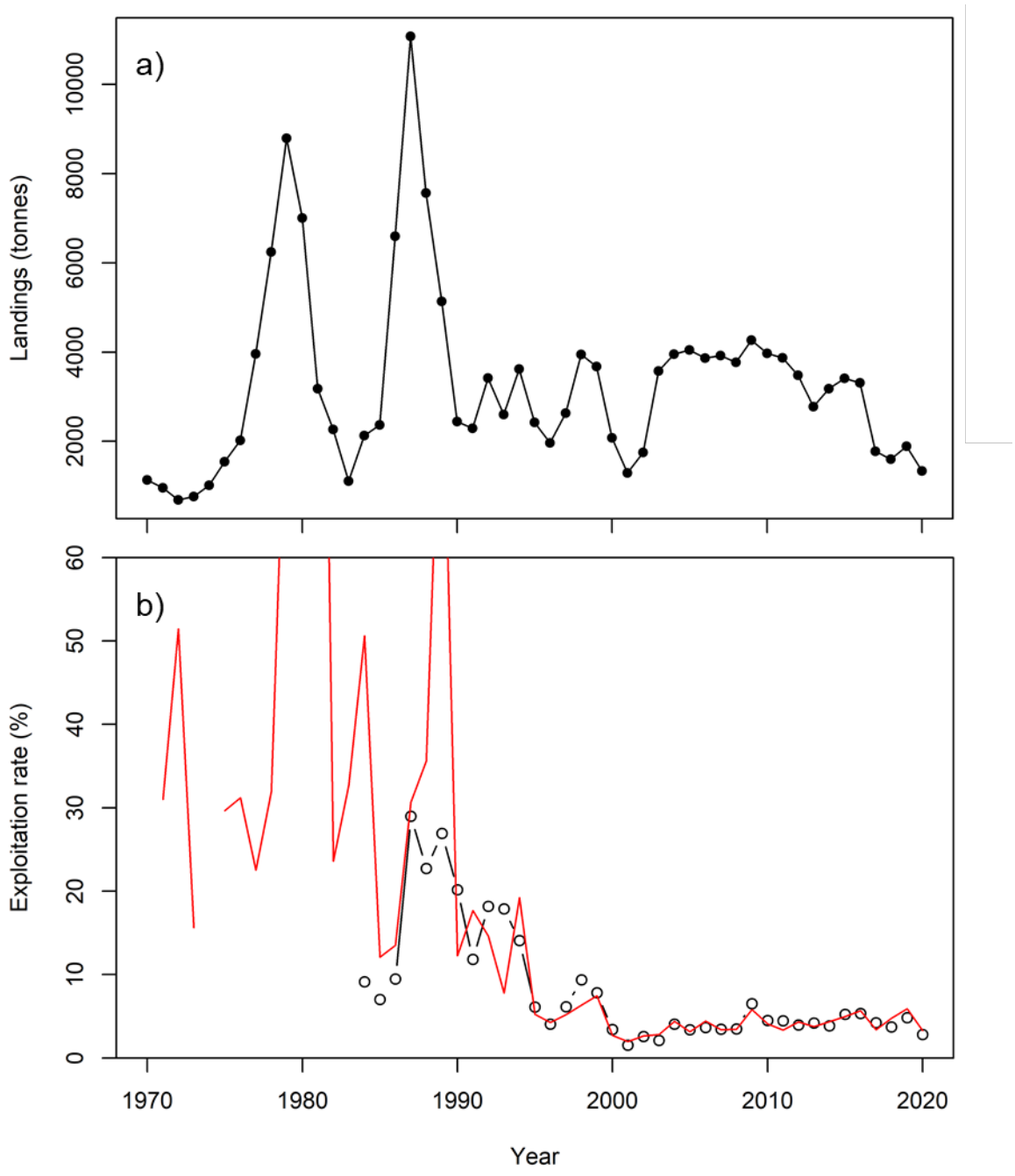

Figure 28. a) Total annual fishery landings of Greenland halibut (tonnes) in 4RST and b) the associated relative exploitation rate (percent) for Greenland halibut >35 cm based on the whole GSL estimated trawlable biomass (black circles and line) and estimate based on the sGSL biomass index and the density dependent relationship with whole GSL trawlable biomass (red line).