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 spatiotemporal overlap

Yihao Yin and Hugues P. Benoît

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

2022

Canadian Technical Report of
Fisheries and Aquatic Sciences 3454

Canadian Technical Report of Fisheries and Aquatic Sciences

Technical reports contain scientific and technical information that contributes to existing knowledge but which is not normally appropriate for primary literature. Technical reports are directed primarily toward a worldwide audience and have an international distribution. No restriction is placed on subject matter and the series reflects the broad interests and policies of Fisheries and Oceans Canada, namely, fisheries and aquatic sciences.

Technical reports may be cited as full publications. The correct citation appears above the abstract of each report. Each report is abstracted in the data base *Aquatic Sciences and Fisheries Abstracts*.

Technical reports are produced regionally but are numbered nationally. Requests for individual reports will be filled by the issuing establishment listed on the front cover and title page.

Numbers 1-456 in this series were issued as Technical Reports of the Fisheries Research Board of Canada. Numbers 457-714 were issued as Department of the Environment, Fisheries and Marine Service, Research and Development Directorate Technical Reports. Numbers 715-924 were issued as Department of Fisheries and Environment, Fisheries and Marine Service Technical Reports. The current series name was changed with report number 925.

Rapport technique canadien des sciences halieutiques et aquatiques

Les rapports techniques contiennent des renseignements scientifiques et techniques qui constituent une contribution aux connaissances actuelles, mais qui ne sont pas normalement appropriés pour la publication dans un journal scientifique. Les rapports techniques sont destinés essentiellement à un public international et ils sont distribués à cet échelon. II n'y a aucune restriction quant au sujet; de fait, la série reflète la vaste gamme des intérêts et des politiques de Pêches et Océans Canada, c'est-à-dire les sciences halieutiques et aquatiques.

Les rapports techniques peuvent être cités comme des publications à part entière. Le titre exact figure au-dessus du résumé de chaque rapport. Les rapports techniques sont résumés dans la base de données *Résumés des sciences aquatiques et halieutiques.*

Les rapports techniques sont produits à l'échelon régional, mais numérotés à l'échelon national. Les demandes de rapports seront satisfaites par l'établissement auteur dont le nom figure sur la couverture et la page du titre.

Les numéros 1 à 456 de cette série ont été publiés à titre de Rapports techniques de l'Office des recherches sur les pêcheries du Canada. Les numéros 457 à 714 sont parus à titre de Rapports techniques de la Direction générale de la recherche et du développement, Service des pêches et de la mer, ministère de l'Environnement. Les numéros 715 à 924 ont été publiés à titre de Rapports techniques du Service des pêches et de la mer, ministère des Pêches et de l'Environnement. Le nom actuel de la série a été établi lors de la parution du numéro 925.

Canadian Technical Report of
Fisheries and Aquatic Sciences 3454

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 SPATIOTEMPORAL OVERLAP

by

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

Cat. No. Fs 97-6/3454E-PDF ISBN 978-0-660-41276-4 ISSN 1488-5379

Correct citation for this publication:

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 spatiotemporal overlap. Can. Tech. Rep. Fish. Aquat. Sci. 3454: xi + 72 p.

TABLE OF CONTENTS

1	INTRODUCTION	1
2	METHOD	2
3	RESULTS AND DISCUSSION	11
4	CONCLUSIONS	17
5	ACKNOWLEDGEMENTS	17
6	REFERENCES	17
7	TABLES AND FIGURES	20

List of 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). Note that exceptionally in 2003, the CCGS Wilfred Templemen was used in the SGSL survey. This vessel is the sister ship to the CCGS Alfred Needler, because these vessels shared the same parameters, the W. Templeman is not included in the table. 20

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). 21

Table 3: Pairs of vessel-gears and number of stations for the two northern Gulf of St. Lawrence comparative fishing experiments (i.e., NGSL 1990 and NGSL 2004-2005) and the three southern Gulf of St. Lawrence comparative fishing experiments (i.e., SGSL 1985, SGSL 1992, and SGSL 2004-2005). 22

Table 4: 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. 23

Table 5: 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..... 24

Table 6: Number of annual survey tows within the overlap area for both northern and southern Gulf of St. Lawrence (NGSL and SGSL). The column headers identify the vessels and trawls. The vessels include E.E. Prince (EE), MV Lady Hammond (LH), CCGS Alfred Needler (AN) and CCGS Teleost (Te). 25

Table 7: Catches of Atlantic halibut from historical comparative fishing (CF) experiments. Number of effective stations is the number of stations where at least one halibut was caught by either gear within the pair. Number of total stations is the total number of successful comparative fishing stations during the survey. Average number of observations is the average number of effective pairs over the total number of 1cm length bins. 27

Table 8: Model selection for the overlap analysis of Atlantic halibut between a smoothing length effect and a conversion without length effect. 28

Table 9: Summary of redfish catch data from the five comparative fishing experiments. Effective stations are where a pair of tows resulted in at least one catch. The ratio estimator is the ratio between the total catches of the old and the new vessels (both effort-standardized). 29

Table 10: Model selection for the five comparative fishing analyses: difference in AIC of all converged models from the best model (indicated by zeros in bold)..... 30

List of Figures

Figure 1: Stratification scheme for the southern Gulf of St. Lawrence multi-species bottom-trawl survey. The annual sampling design has included strata 415-439 since 1971 and strata 401-403 since 1984. 31

Figure 2: Stratification scheme for the northern Gulf of St. Lawrence multi-species bottom-trawl survey. The annual sampling design has included strata 401-408, 801-824 and 830 since 1984, and 827-829, 831 and 832 since 1985. Additional strata, located in NAFO area 3Ps (southwest Newfoundland) and sampled only in 1987 and 1993-2003 are not shown. 32

Figure 3: 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. . 33

Figure 4: Survey tow locations for the five comparative fishing experiments, two in the northern Gulf of St. Lawrence (NGSL) and three in the southern Gulf of St. Lawrence (SGSL). 34

Figure 5: Tow locations within the overlap area from both the northern and southern Gulf of St. Lawrence (NGSL and SGSL) RV surveys and the scheme of their stratification into three substrata for the overlap analysis (colored points). Blue lines border the six NGSL RV survey strata within the area, red lines border the three SGSL RV survey strata; only survey tows within both surveys were included for the overlap analysis. 35

Figure 6: Total catches of Atlantic halibut by each vessel-gear from the annual RV surveys in both southern and northern Gulf of St. Lawrence (catch numbers were effort-adjusted according to the standard in the respective surveys). 36

Figure 7: Latitude distribution of catches by *Teleost*-Campelen and *Teleost*-WIIA in the overlap area for four length groups of Atlantic halibut. Note that because of the shape and arrangement of the area, latitude acts as a proxy for location along the three substrata (see Figures 3 and 5). 37

Figure 8: Catch-length composition for Atlantic halibut from the annual surveys in the northern and southern Gulf of St. Lawrence (NGSL and SGSL) in the overlap area. Catches were aggregated every 5cm between 30-110cm. 38

Figure 9: Annual trend of mean catch per tow for four length groups of Atlantic halibut using catches within the modeled range, within 2006 to 2020, and from the overlap area. 39

Figure 10: Estimated relative catch efficiencies between *Teleost*-WIIA and *Teleost*-Campelen on the log scale with one standard deviation by the two candidate models BI0 and BI2 in the overlap analysis of Atlantic halibut. 40

Figure 11: *DHARMa* residual diagnostics for the catch distribution assumption, including tests for zero-inflation (top-left panel), over-dispersion (top-right panel) and goodness-of-fit of the negative binomial distribution assumption (bottom-left panel), as outputted from the DHARMa package for R. The model past all tests, indicating a suitable fit. 41

Figure 12: Catch of redfish (standardized number) by paired vessels from the comparative fishing experiment in the northern Gulf of St. Lawrence in 1990 in four length groups (solid red circles for positive catches and open circles for zero catches)..... 42

Figure 13: Catch of redfish (standardized number) by paired vessels from the comparative fishing experiment in the northern Gulf of St. Lawrence in 2004-2005 in four length groups (solid red circles for positive catches and open circles for zero catches). 43

Figure 14: Catch of redfish (standardized number) by paired vessels from the comparative fishing experiment in the southern Gulf of St. Lawrence in 1985 in four length groups (solid red circles for positive catches and open circles for zero catches)..... 44

Figure 15: Catch of redfish (standardized number) by paired vessels from the comparative fishing experiment in the southern Gulf of St. Lawrence in 1992 in four length groups (solid red circles for positive catches and open circles for zero catches)..... 45

Figure 16: Catch of redfish (standardized number) by paired vessels from the comparative fishing experiment in the southern Gulf of St. Lawrence in 2004-2005 in four length groups (solid red circles for positive catches and open circles for zero catches). 46

Figure 17: Comparative fishing analysis of NGSL 1990, between Lady Hammond-WIIA and *Alfred Needler*-URI: Estimated proportion of catch over length by 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). 47

Figure 18: Comparative fishing analysis of NGSL 2004-2005, between *Alfred Needler*-URI and *Teleost*-Campelen: Estimated proportion of catch over length by *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), 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)..... 48

Figure 19: Comparative fishing analysis of SGSL 1985, between EE Prince-Yankee and Lady Hammond-WIIA: Estimated proportion of catch over length by 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)..... 49

Figure 20: Comparative fishing analysis of SGSL 1992, between Lady Hammond-WIIA and *Alfred Needler*-WIIA: Estimated proportion of catch over length by 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)..... 50

Figure 21: Comparative fishing analysis of SGSL 2004-2005, between *Alfred Needler*-WIIA and *Teleost*-WIIA: Estimated proportion of catch over length by *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). 51

Figure 22: Estimated relative catch efficiency as a function of length from each comparative fishing analysis (black line), with one standard deviation (grey band). The red line represents a relative catch efficiency of 1, indicating no difference between the pair of vessel-gears. 52

Figure 23: Comparative fishing analysis of NGSL 1990, between Lady Hammond-WIIA and *Alfred Needler*-URI: normalized randomized quantile residuals for each station (top panel) and for each length bin (bottom panel). The boxes indicate the 25% and 75% quantiles and the segment is the median; whiskers extends from the hinge to the largest/smallest values no further than 1.5 * IQR from the hinge (where IQR is the distance between the first and third quartiles). 53

Figure 24: Comparative fishing analysis of NGSL 2004-2005, between *Alfred Needler*-URI and *Teleost*-Campelen: normalized randomized quantile residuals for each station (top panel) and for each length bin (bottom panel). The boxes indicate the 25% and 75% quantiles and the segment is the median; whiskers extends from the hinge to the largest/smallest values no further than 1.5 * IQR from the hinge (where IQR is the distance between the first and third quartiles). 54

Figure 25: Comparative fishing analysis of SGSL 1985, between EE Prince-Yankee and Lady Hammond-WIIA: normalized randomized quantile residuals for each station (top panel) and for each length bin (bottom panel). The boxes indicate the 25% and 75% quantiles and the segment is the median; whiskers extends from the hinge to the largest/smallest values no further than 1.5 * IQR from the hinge (where IQR is the distance between the first and third quartiles). 55

Figure 26: Comparative fishing analysis of SGSL 1992, between Lady Hammond-WIIA and *Alfred Needler*-WIIA: normalized randomized quantile residuals for each station (top panel) and for each length bin (bottom panel). The boxes indicate the 25% and 75% quantiles and the segment is the median; whiskers extends from the hinge to the largest/smallest values no further than 1.5 * IQR from the hinge (where IQR is the distance between the first and third quartiles). 56

Figure 27: Comparative fishing analysis of SGSL 2004-2005, between *Alfred Needler*-WIIA and *Teleost*-WIIA: normalized randomized quantile residuals for each station (top panel) and for each length bin (bottom panel). The boxes indicate the 25% and 75% quantiles and the segment is the median; whiskers extends from the hinge to the largest/smallest values no further than 1.5 * IQR from the hinge (where IQR is the distance between the first and third quartiles). 57

Figure 28: Annual total catch of redfish (effort-standardized catch numbers) by each vessel and gear using RV survey data from both northern and southern Gulf of St. Lawrence in the area of overlap. Catch levels have surged since 2013 as indicated by *Teleost*-Campelen. 58

Figure 29: Time series of mean catch per tow for each vessel-gear and for four length groups using survey catches in the overlap area between the northern and southern Gulf of St. Lawrence (NGSL and SGSL). 59

Figure 30: Spatial distribution of redfish survey catches (effort-adjusted) in the overlap area by each vessel-gear during 1984-1992 for a comparison between E.E. Prince(EE)-Yankee and Lady Hammond(LH)-WIIA in both the northern Gulf of St. Lawrence (denoted by N in the figure) and southern Gulf of St. Lawrence (denoted by S). 60

Figure 31: Spatial distribution of redfish survey catches (effort-adjusted) in the overlap area by each vessel-gear during 1990-2005 for a comparison between *Alfred Needler*(AN)-WIIA in the southern Gulf of St. Lawrence and *Alfred Needler*-URI in the northern Gulf of St. Lawrence. Note that 2003-AN|WIIA was in fact undertaken by the *CCGS W. Templemen*, sister ship to the *CCGS A. Needler*. 61

Figure 32: Spatial distribution of redfish survey catches (effort-adjusted) in the overlap area by each vessel-gear during 2004-2020 for a comparison between *Teleos*t(Te)-WIIA in the southern Gulf of St. Lawrence and Teleost-Campelen in the northern Gulf of St. Lawrence. 62

Figure 33: Estimated log-relative catch efficiency for each gear (all gears calibrated to Teleost-Campelen equivalent, in black horizontal line). 63

Figure 34: Estimated catch efficiencies with one standard deviation estimated from the integrated model and transformed to be pairwise corresponding to the five comparative fishing experiments, in comparison with results from the separate comparative fishing (CF) analyses. Estimates are plotted in red and blue lines for the two models, respectively, and the bands are standard deviations. 64

Figure 35: Residual diagnostics for paired catches: normalized randomized quantile residuals were calculated for each pair and each length bin based on the beta-binomial distribution and are generally compared to a gaussian distribution. 65

Figure 36: Residual diagnostics for paired catches: randomized quantile residuals were checked for each comparative fishing experiment and each length bin (black dots, jittered to enhance clarity). 66

Figure 37: *DHARMa* residual diagnostics for the catch distribution assumption in the integrated analysis, including tests for zero-inflation, over-dispersion and goodness-of-fit of the negative binomial distribution. The model past all tests except the KS-test, indicating some deviation from the overall distributional assumption. 67

Figure 38: Predicted versus observed catches for each 2cm length group (any quantities under 10 were conformed to 10 exactly in order to preserve all data as well as the log scale

of the axes in the plot). The name of each panel indicates the median of the length group, e.g., 5.5 indicates the length group 5-6cm. 68

Figure 39: Annual trend (upper panel) and length relationship (lower panel) of the estimated zero-inflation rate, or proportion of true zeros in the overlap area. 69

Figure 40: Sample proportion of zero catches (crosses) for the three length groups chosen for illustration: 11-12cm, 21-22cm, and 31-32cm (from top to bottom), and their corresponding estimated zero-inflation rate (in circles), or proportion of true zeros. The zero-inflation rate (zip) is an indicator of redfish presence probability (1-zip) within the area. 70

Figure 41: Standardized redfish abundance indices (mean numbers per tow, with 95% confidence interval) for the entire GSL with standardization based on comparative fishing only (black points) and from the integrated analysis (grey points) for all sizes and by size groups used in the assessment of Units 1+2 redfish. 71

Figure 42: Annual relative length frequencies (proportions at length) for the standardized redfish abundance indices for the entire GSL, with standardization based on comparative fishing only (black lines) and from the integrated analysis (blue lines). 72

ABSTRACT

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 spatiotemporal overlap. Can. Tech. Rep. Fish. Aquat. Sci. 3454: xi + 72 p.

Standardized bottom-trawl surveys provide fishery-independent information on relative abundance for the assessment and management of demersal fish stocks worldwide. These surveys follow standard procedures/protocols to maintain a consistent catchability to avoid confounding actual changes in relative abundance. In the Gulf of St. Lawrence (GSL), two annual research vessel surveys have been conducted, one in the northern GSL and Estuary (NGSL) and one in the southern GSL (SGSL). These surveys have employed different vessels and gears except during 1984-1990. Each time the vessel or gear changed, a paired-tow comparative fishing experiment was conducted to estimate relative catch efficiency between the former and replacement vessels and gears. Given that the two surveys employed the same vessel and gear in some years in the 1980s, jointly standardizing the surveys is possible by sequentially applying the results of comparative fishing. In addition, both surveys have, with few exceptions, consistently sampled the same area along the southern slope of the Laurentian channel annually. Catches from this area of spatial overlap can also inform the relative catch efficiency of the two surveys. Here, we develop an overlap analysis based on catches in the overlap area of the NGSL and SGSL surveys and apply to Atlantic Halibut (Hippoglossus hippoglossus) to calibrate and combine the relative abundance indices of the two surveys for 2006-2020. Standardisation for prior years was not possible because Atlantic Halibut was less abundant and too few were caught during comparative fishing or in the area of spatial overlap to produce reliable estimates of relative efficiency. We also re-analyze the five comparative fishing experiments conducted in both NGSL and SGSL during 1984-2020 for redfish (Sebastes spp.) and then integrate the overlap analysis with these comparative fishing analyses to inter-calibrate all vessels and gears in both surveys to develop a combined GSL-wide survey index.

RÉSUMÉ

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 spatiotemporal overlap. Can. Tech. Rep. Fish. Aquat. Sci. 3454: xi + 72 p.

Les relevés standardisés au chalut de fond fournissent partout au monde des informations indépendantes de la pêche sur l'abondance relative pour l'évaluation et la gestion des stocks de poissons démersaux. Ces relevés suivent des procédures/protocoles standard pour maintenir une capturabilité cohérente afin d'éviter toute confusion avec les changements réels de l'abondance relative. Dans le golfe du Saint-Laurent (GSL), deux relevés annuels par navire de recherche ont été effectués, un dans le nord du GSL et de l'estuaire (NGSL) et un dans le sud du GSL (SGSL). Ces relevés ont utilisé des navires et des engins différents, sauf entre 1984 et 1990. Chaque fois que le navire ou l'engin changeait, une expérience de pêche comparative par traits appariés était menée pour estimer l'efficacité relative de capture entre les anciens navires et engins de remplacement. Étant donné que les deux relevés utilisaient le même navire et le même engin certaines années dans les années 1980, il est possible de normaliser conjointement les deux relevés en appliquant séquentiellement les résultats de la pêche comparative. De plus, les deux relevés ont, à quelques exceptions près, échantillonné systématiquement la même zone le long du versant sud du chenal Laurentien chaque année. Les captures de cette zone de chevauchement spatial peuvent également renseigner sur l'efficacité relative des captures des deux relevés. Ici, nous développons une analyse de chevauchement basée sur les captures dans la zone de chevauchement des relevés NGSL et SGSL et l'appliquons au flétan atlantique (*Hippoglossus hippoglossus*) pour calibrer et combiner les indices d'abondance relative des deux relevés pour 2006- 2020. La normalisation pour les années précédentes n'a pas été possible parce que le flétan de l'Atlantique était moins abondant et trop peu ont été capturés lors de la pêche comparative ou dans la zone de chevauchement spatial pour produire des estimations fiables de l'efficacité relative. Nous analysons également à nouveau les cinq expériences de pêche comparative menées dans le NGSL et le SGSL entre 1984 et 2020 pour le sébaste (Sebastes spp.). Par la suite nous intégrons ces analyses à celles pour la zone de chevauchement afin de développer un indice d'abondance combinant les deux relevés à l'échelle du GSL.

1 INTRODUCTION

Standardized bottom-trawl surveys are an important source of fishery-independent information on relative abundance that is key to the assessment and management of demersal fish stocks worldwide. Survey procedures and protocols are standardized to avoid introducing systematic changes in catchability that could otherwise be confounded with actual changes in abundance. For instance, calibration or standardization is employed when it is necessary to replace ageing vessels, or to change aspects such as the sampling gear or other protocols to better match survey objectives (Bagley et al., 2015). Experiments based on paired-tow comparative fishing are the gold standard for estimating and accounting for changes in relative catchability caused by such structural changes in the surveys (e.g., Pelletier, 1998; Lewy et al., 2004; Cadigan and Dowden, 2010; Miller, 2013). Such experiments have been undertaken in all major bottom-trawl research surveys in Atlantic Canada since at least the early 1980s each time a vessel or trawl was changed (e.g., Warren, 1997; Benoît, 2006; Bourdages et al., 2007; Fowler and Showell, 2009). The experimental design of paired-tow surveys is considered efficient for estimating the relative catchability of survey vessels and gear because it accounts for medium to large scale spatial-temporal variation in fish density and in factors that can affect catchability such as depth and bottom-type. However, other statistical modelling approaches based on spatiotemporal matching at somewhat larger scales have also been employed when comparative fishing is not feasible (e.g., Thorson and Ward, 2014). This was the case in the Sentinel bottom-trawl survey of the southern Gulf of St. Lawrence, which involved several vessels each year, with some changes across years (Savoie, 2014).

Two research vessel surveys are conducted annually in the Gulf of St. Lawrence (GSL), one in the southern GSL (SGSL survey) and one in the northern GSL and Estuary (NGSL survey) (Figures 1 and 2). Each has employed different vessels and bottom trawls (gears) over their history. The SGSL survey has been conducted 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 MV Lady Hammond (1985-1992), the the Canadian Coast Guard Ship (CCGS) *Alfred Needler* (1992-2005) and the CCGS *Teleost* (2004-present). Exceptionally in 2003 the CCGS *Wilfred Templeman*, sister ship to the CCGS *Alfred Needler*, fishing the WIIA trawl, was used for the SGSL survey. 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 (Benoît and Swain, 2003; Benoît 2006). The results of these experiments have been analyzed and applied routinely to maintain the integrity of the standardized abundance series for a large number of taxa.

The NGSL survey has been conducted each August since 1984, first by the Lady Hammond fishing a Western IIA trawl (1984-1990), subsequently by the CCGS *Alfred Needler* fishing a URI trawl (1990-2005) and then by the CCGS *Teleost* fishing a Campelen 1800 trawl (2004-present). Comparative fishing experiments involving former and replacement vessels and trawls were conducted during the regular surveys in 1990, 2004 and 2005 (Gascon et al., 1991; Bourdages et al., 2007). While the results of the 2004-2005 experiments have previously been analyzed for a large number of taxa (Bourdages et al., 2007), the results of the 1990 experiment have only been analyzed for witch flounder (*Glyptocephalus cynoglossus*) (Swain et al., 1998), Greenland halibut (*Reinhardtius hippoglossoides*) (Yin and Benoît, 2022), Atlantic cod (*Gadus morhua*) (Benoît et al., in review) and redfish

(*Sebastes spp.*), although the analyses for redfish have not been peer-reviewed and the details are not published.

Species that occur in the deeper waters of the GSL (depths below 150 m) are typically captured in both the SGSL and NGSL surveys. These species include witch flounder, Greenland halibut, white hake (*Urophycis tenuis*), Atlantic halibut (*Hippoglossus hippoglossus*), and redfish. To estimate abundance indices that cover the distributional range of these species, or species complexes in the case of redfish, in the GSL, data from these two surveys should ideally be combined. Although the two surveys share a common vessel-gear tandem used in the 1980s (Lady Hammond-WIIA) that would, in principle, permit joint standardization of the two surveys over the time series, this has only been undertaken for witch flounder and Greenland halibut (Swain et al., 1998; Yin and Benoît, 2022).

In addition to a shared predecessor vessel and gear, the two surveys also overlap spatially along the southern slope of the Laurentian channel, and both surveys take place in the same season, albeit with a one month separation (the median separation over the history of the surveys is 27 days). Provided there are no seasonal movements of fish into or out of this area during the August-September period, this spatial overlap provides information on the relative catchability of the two surveys based on annual catches in the area using approaches such as that of Thorson and Ward (2014). In principle, this should improve estimates of relative catchability of the two surveys if that analysis is integrated with a treatment of the comparative fishing data. It may also improve survey-specific estimates for the change from the Lady Hammond to the CCGS *Alfred Needler*, because these changes did not occur in the same year in both surveys; catches in the survey in which the change was not occurring that year can thus effectively serve as a baseline against which to measure relative catch efficiency changes in the other survey. For some stocks such as Atlantic Halibut, for which abundance was previously too low to generate sufficient catches to estimate relative catchability using comparative fishing results, standardization of the two surveys to one another is only possible using data from the overlapping area.

In this document we re-analyse the results of comparative fishing experiments from both surveys for Atlantic halibut and redfish. We apply statistical methods that have improved flexibility and can better account for different sources of error (Miller 2013) compared to the methods employed in the initial analysis of the data (Benoît 2006; Bourdages et al., 2007). We also model catches using spatiotemporal matching to estimate relative catch efficiency, and apply these models in both a stand-alone analysis and an analysis that integrates the paired-tow data. We then evaluate the potential to generate standardized GSL-wide abundance indices for the two species that begin in 1984. The stand-alone and integrated spatiotemporal analyses are a novel aspect of this and recently completed work (Yin and Benoît, 2022). These methods should be broadly applicable to the calibration of other surveys that overlap spatially and that can be assumed to sample the same underlying densities of fish.

2 METHOD

2.1 Data

The data for this analysis all result from standardized survey sampling, mostly in the context of regular survey operations that in some cases also involved paired-tow comparative

fishing (hereafter, simply 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. The CCGS *Wilfred Templeman* was considered to have equivalent relative-efficiency to its sister ship to the CCGS *Alfred Needler* for the four sets made in the SGSL survey in 2003 which were relevant for our analyses.

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. The strata that comprise the area of overlap between the two surveys are 415, 425, and 439 in the SGSL survey, and 401-406 in the NGSL survey (Figure 3). 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 comparative fishing experiments treated in our analyses are available in other reports and are not repeated here, with the exception of presenting sample sizes and showing the location of the paired fishing sets (Figure 4). Readers are referred to the following documents for details: Benoît and Swain (2003) for comparative fishing in the SGSL in 1995 and 1992; Benoît (2006) for the SGSL in 2004 and 2005; Yin and Benoît (2022) for comparative fishing in the NGSL in 1990; and, Bourdages et al. (2007) for the comparative fishing in the NGSL in 2004 and 2005.

In what follows, we use the term *station* to denote the geographic location selected for fishing by one vessel during regular survey operations and by the pair of vessels during comparative fishing. The terms *set* and *tow* refer to a single fishing event. Thus, during comparative fishing, two vessels each undertake a set/tow as close as is practical and safe at the same station. *Catch* refers to the capture that results from a set/tow.

2.2 Comparative Fishing Analysis

In this section, a suite of Binomial and Beta-Binomials models are presented for the analysis of paired catches from comparative fishing experiments. To improve data quality and model estimability in the analysis, catches are aggregated by 1cm length bins although lengths are measured to the millimeter during the NGSL surveys.

2.2.1 Binomial Models

To estimate the relative catch efficiency between paired gears (“gear” in this section refers to a vessel-gear combination) in the comparative fishing, we assume the expected catch from gear g ($g \in \{A, B\}$) at length l and at station i is,

$$E\left[C_{gi}(l)\right] = q_{gi}(l) D_{gi}(l) f_{gi},$$

where $q_{gi}(l)$ is the catchability of gear g, as a function of fish length l, D_{gi} is the underlying population density sampled by gear g, and f_{gi} is a standardization term which usually

includes the swept area of a tow and if applicable, the proportion of sub-sampling for size measurement on-board. In a binomial model, catch from gear A at station i, conditioning on the combined catch from both gears within this station, $C_i(l) = C_{Ai}(l) + C_{Bi}(l)$, is binomial-distributed,

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

where $p_{Ai}(l)$ is the expected proportion of catch from gear A. Paired tows are assumed to sample the same underlying density, $D_{Ai}(l) = D_{Bi}(l) = D_i(l)$, hence the logit-probability of catch by gear A is:

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

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

This gives the conversion factor, $\rho_i(l)$, as the ratio of catchabilities between gear A and B at length l and at station i,

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

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 (Wood, 2006), 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, and the basis functions and penalty matrices were generate by the *R* package *mgcv* (Wood, 2011).

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,

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

where $\mathcal{L}$ is 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 functions, and λ is the smoothness parameter. This smoothness parameter is estimated by maximum likelihood along with other model parameters but may be sensitive to the data and in such cases, can be determined by other criteria such as generalized cross-validation.

The penalized maximum likelihood smoother can also be re-parameterized into a mixed effects model,

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

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 $\boldsymbol{D}_+$ denotes the diagonal matrix of the positive eigenvalues of $\mathbf{S}$. 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 bounded by $K-2$. 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 among stations,

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

where $\boldsymbol{\delta}_i \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$ and $\boldsymbol{\epsilon}_i \sim \mathrm{N}(\mathbf{0}, \mathbf{D}_+^{-1}/\xi)$. From a similar re-parameterization of the cubic spline smoother, these random effects allow for deviations of the length-based conversion at each station. $\boldsymbol{\Sigma}$ is the covariance matrix of the random effects corresponding to the random deviations and contains 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,

$$\begin{cases} C_{Ai}(l) \sim BI(C_i(l), p_{Ai}(l)) \\ C_i(l) = C_{Ai}(l) + C_{Bi}(l) \\ logit(p_{Ai}(l)) = log(\rho_i(l)) + o_i \\ log(\rho_i(l)) = \mathbf{X}_f^T(\boldsymbol{\beta}_f + \boldsymbol{\delta}_i) + \mathbf{X}_r^T(\mathbf{b} + \boldsymbol{\epsilon}_i) \end{cases}$$

The model is estimated via maximum likelihood and the marginal likelihood integrating out random effects is,

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

The binomial mixed model can take various assumptions on the smoother and station variation to accommodate different underlying density of a species and data limitations especially in length measurements. A set of binomial models, based on subsets of the full model are presented in Table 4.

2.2.2 Beta-Binomial Models

The binomial assumption for the catch can be extended to a beta-binomial distribution to allow over-dispersion at the stations,

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

The beta-binomial distribution is a compound of the binomial distribution and a beta distribution. More specifically, it assumes a beta-distributed random effect in the expected proportion of catch from gear A across stations. As a result, the expected catch by gear *A* has a variance of

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

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

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

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

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.

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

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

The marginal likelihood is,

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

Likewise, various smoothing assumptions can be applied to the variance parameter. Table 5 presents a set of beta-binomial mixed models.

2.2.3 Model Selection

In the analysis of each comparative fishing experiment, the conversion factor was developed for a pre-specified length range at an interval of 1cm. The length range was usually selected to be from the minimum to the maximum observed length during the comparative fishing survey, but may be narrowed to exclude extremely large or small individuals in sporadic catches to avoid disproportional impact from outliers.

The binomial and beta-binomial models in Tables 4 and 5 were implemented in Template Model Builder (TMB, Kristensen et al, 2016) in which they 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 equally-spaced knots ($K = 9$) within the pre-specified length range depending on the comparative fishing survey. TMB automatically calculates a standard error for the maximum likelihood estimation of the conversion factor via the delta method (Kristensen et al., 2016).

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 Akaike information criterion (AIC) 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 comprehensive 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, if available.

2.2.4 Survey Index Calibration

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:

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

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

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

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

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

In this report, we proceed with calibrations as they have traditionally been employed, that is without propagating their uncertainty to the estimated uncertainty in catch related estimates, such as abundance indices. Propagation of uncertainty could be developed by using computer intensive approaches such as bootstrapping.

Length-dependent relative catch efficiency was estimated only over the range of lengths 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 **Survey Overlap Analysis**

2.3.1 Negative Binomial Model

The spatial overlap between the southern Gulf of St. Lawrence (SGSL) and northern Gulf of St. Lawrence (NGSL) annual RV surveys has resulted in sample tows covering the same area consistently for over three decades (Figure 4, Table 6). This intersection of sampling coverage can provide information for the relative catch efficiency between different gears deployed by the two separate surveys based on their shared underlying population density. Despite multiple gear updates in both survey regions over time, most gears have been used for multiple years and have generated some amount of effective survey tows within the overlap area (Table 6), especially for species that are abundant within the area.

In the survey overlap analysis, the area of overlap between the two surveys was divided into three sub-areas utilizing the SGSL stratification scheme (Figure 5) for a spatial aggregation of survey tows, hereafter referred to as "substrata" to distinguish from "strata" in the RV surveys. Similar to design-based stratified RV survey analysis, we assume homogeneous population density, $D_{s,t}$, within each substratum $s \in \{1,2,3\}$ and year t. The substrata were used to allow reasonable spatial variation within the area and to potentially increase estimation accuracy. The expected catch by gear g at station i and at length l is

$$\mu_{git}(l) = E\left[C_{git}(l)\right] = q_g(l) D_{s_i,t}(l) f_{gi},$$

where $q_g(l)$ is the catchability of gear g, $D_{s_i,t}$ is the underlying density at station i in year t, and f_{gi} is the standardization term associated with each tow including swept area and proportion of sub-sampling for length measurement.

The objective of this study is to calibrate all gears to *Teleost*-Campelen equivalent catches. Therefore, the quantity of interest is the relative catch efficiency between any gear g and the standard gear g_0 (*Teleost*-Campelen), $\rho_g(l)$. This gives

$$\mu_{git}(l) = \rho_g(l) \cdot q_{g_0}(l) D_{s_i,t}(l) f_{gi} = \rho_g(l) \mu_{0it}(l) o_{gi},$$

where $o_{gi} = f_{gi}/f_{0i}$ is an offset term between gears g and g_0 and $\mu_{0it}(l)$ is the expected catch calibrated to g_0 (density as "seen" by the standard gear g_0) which can be estimated along with the conversion factors. The analysis is focused on deriving viable calibrations among gears and in the "bias-variance" trade-off, should favor minimal estimation bias. For this reason, the densities are estimated as fixed effects without additional structural constraints that may add to bias.

Survey catch numbers at length are assumed to follow a zero-inflated negative binomial (ZINB) distribution within each length bin, year, and substratum,

$$C_{git}(l) \sim ZINB\left(\mu_{git}(l), \alpha, \pi\right),$$

where α and π are the over-dispersion and zero-inflation parameters, respectively. The probability mass function (PMF) of ZINB is as follows,

$$p_{ZINB}[C = j] = \begin{cases} \pi + (1-\pi) \cdot p_{NB}[C = 0], & if \quad j = 0 \\ (1-\pi) \cdot p_{NB}[C = j], & if \quad j > 0 \end{cases}$$

where p_{NB} is the PMF of the corresponding negative binomial distribution, $NB(\mu_{git}(l), \alpha)$,

$$p_{NB}[C = j] = \frac{\Gamma(j + \alpha^{-1})}{\Gamma(j+1)\Gamma(\alpha^{-1})} \left(\frac{1}{1+\alpha\mu}\right)^{\alpha^{-1}} \left(\frac{\alpha\mu}{1+\alpha\mu}\right)^{j}.$$

The over-dispersion considers extra-Poisson variation and the zero-inflation represents presence-absence. These two parameters are critical to the modeling of species abundance and can also accommodate fine scale spatial distributions or potential distributional shifts to a certain extent, sparing an explicit spatiotemporal modeling of underlying density that in practice, is both complex and highly reliant on data sufficiency. In the analysis, the ZINB may be reduced to a Negative Binomial (NB), Zero-Inflated Poisson (ZIPois), or simply, Poisson (Pois) distribution, and parameter(s) of the distribution may be extended to allow

variation according to substratum, year, length, or a combination of these, depending on model convergence and suitability.

Catchability of gear g, and hence its relative catch efficiency to g_0, is assumed as a continuous smooth function of length. The same smoothing technique was used as in the comparative fishing analysis, i.e., a cubic smoothing with re-parameterization into a mixed effects model implemented in TMB:

$$log\left(\rho_g(l)\right) = \mathbf{X}_f^T\boldsymbol{\beta}_f + \mathbf{X}_r^T\mathbf{b}.$$

Model parameters include a smoothness parameter λ, fixed parameters $\boldsymbol{\beta}_f$ and random parameters $\mathbf{b}$ (see equations in section above). The station effect in the binomial and beta-binomial models for the comparative fishing analysis pertaining to potential spatial or random variations during the tows have been assimilated into the ZINB distribution. The station random effects are not applicable, thus candidate models only include BI0 and BI2 for a length-independent and a length-dependent conversion factor respectively. Both models were fitted to the data and the better model was selected based on a lower AIC.

A summary of the above negative binomial model is as follows, where underlying density is sampled by multiple gears to provide information for their relative catch efficiencies,

$$\begin{cases} C_{git}(l) \sim ZINB(\mu_{git}(l), \alpha, \pi) \\ \mu_{git}(l) = \rho_g(l)\mu_{0it}(l)o_{gi} \\ log(\rho_i(l)) = \mathbf{X}_f^T\boldsymbol{\beta}_f + \mathbf{X}_r^T\mathbf{b} \end{cases}$$

The model is estimated via maximum likelihood. The marginal likelihood derived from all samples $\mathbf{Y}_i$ and integrating out random parameters is as follows,

$$\mathcal{L}\left(\boldsymbol{\beta}_f, \lambda, \boldsymbol{\mu}_0, \alpha, \pi\right) = \int \prod_{i=1}^{m} f\left(\mathbf{Y}_i \middle| \mathbf{X}_f, \mathbf{X}_r, \boldsymbol{\beta}_f, \mathbf{b}, \boldsymbol{\mu}_0, \alpha, \pi\right) f(\mathbf{b}|\lambda)\mathrm{d}\mathbf{b}.$$

2.3.2 Integrated Model

The comparative fishing and the survey overlap analyses can be integrated to inter-calibrate relative catch efficiencies among multiple gears that have been deployed in both SGSL and NGSL surveys. The integrated analysis utilizes available information from both sources. The comparative fishing tows from controlled experiments generally produce more accurate calibrations but are limited to specific pairs of gears and are limited to a single experiment. In contrast, the annual survey tows can provide comparisons between any pairs of gears depending on their temporal and spatial overlap but sample size (number of tows sampled by a gear within the overlap area in each year) is usually small and estimation quality is also highly subject to population distributional patterns within the overlap area; however, because the data typically span multiple years, there may be a decreased probability of incorrect estimates compared to controlled experiments done once under the conditions that happened to prevail at the time. Integrating the two types of data and methods is expected to improve estimation quality of these relative catch efficiencies, especially for gears indirectly calibrated to CCGS *Teleost*-Camplen via the multiplicative method (Section 2.2.4). In particular, converting CCGS *Teleost*-WIIA to CCGS *Teleost*-Campelen requires four steps of multiplication where bias and uncertainty may accumulate and overwhelm the estimate.

For the integrated analysis, data included all comparative fishing sets from both SGSL and NGSL (five comparative fishing experiments, Table 3) and annual survey sets within the overlap area (Table 6). The analysis is focused on gear calibration rather than modeling of underlying population density. For this reason, catches outside the overlap area and not part of any comparative fishing experiments were excluded from the analysis to reduce model complexity. Maximum likelihood estimation evaluates the marginal likelihood derived from the joint likelihood of the two components: the binomial (or beta-binomial) model and the negative binomial model. Notably, paired tows within the same stations resulting from comparative fishing were evaluated for the underlying density using their combined catches in order to avoid duplicate contribution to the joint likelihood. Same as the overlap analysis, the over-dispersion and zero-inflation parameters may take various forms according to substratum, year and length. For simplicity, a suitable binomial or beta-binomial model within the 13 candidates was selected for the integrated analysis and fitted to all five experiments based on an assessment of model fitting results from their separate comparative fishing analyses. A comprehensive model selection process would otherwise require a comparison of 13^5 combinations when integrating the five analyses each including 13 candidate models, and is not feasible in practice.

2.3.3 Calibration of Survey Catch

In the overlap analysis and integrated analysis, all gears were calibrated to a selected standard gear g_0 (usually the newest gear). This facilitated the calibration of survey catch into one standard and circumvented a chain of multiplications. Propagation of uncertainty was also straightforward and an estimation error was reported for the catch efficiency of each gear relative to the standard gear. To verify the conversion factors, results from the integrated analysis could be compared to the sequential application of calibration factors from the comparative fishing analyses, or vice versa, by transforming results from the integrated analysis into conversions between the five pairs of vessels and then comparing to results from the comparative fishing analyses directly, as propagating estimation uncertainty is more straightforward in the integrated model.

2.4 **Model Diagnostics**

For model diagnostics of the comparative fishing analysis, residuals were calculated and summarized for catches from one gear (i.e., observed catch numbers by gear A minus model-predicted catches for gear A at each station) as the binomial and beta binomial models were designed to fit the catch distribution by one gear within a pair of catches. In order to assess residual bias against each length bin and each station, the normalized and randomized quantile residuals (Dunn and Smyth, 1996) were derived and presented for each analysis.

The overlap analysis and the integrated analysis both fit negative binomial distributions to the survey catches and a rigorous model diagnostics regime provided by the *DHARMa R* package (Hartig, 2020) was used to evaluate model fit in these cases. *DHARMa* uses a simulation-based approach to create scaled quantile residuals for mixed effects models so that they can be interpreted in a manner similar to simple regression models. The package generates simulated samples based on estimated parameters from the fitted model and tests simulated samples against original data to assess goodness-of-fit of the model. These statistical tests include a test for general uniformity, where *DHARMa* applies the Kolmogorov–Smirnov (KS) test (Marsaglia et al., 2003) and compares the empirical

cumulative distribution functions of the simulated samples against the negative binomial distribution to detect deviation from the assumed distribution, dispersion, and outliers. In addition, there are tests for zero-inflation and over-dispersion, which are especially useful for the discrete negative binomial distribution assumption for catch numbers.

3 RESULTS AND DISCUSSION

3.1 Atlantic Halibut: Overlap Analysis

Due to low population levels in the GSL and possible low catch efficiency of the gears, the five comparative fishing experiments in both SGSL and NGSL did not result in sufficient catches of Atlantic halibut for comparative fishing analysis (Table 7). Previous analyses for the change from the CCGS Needler to CCGS Teleost had concluded no change in efficiency in either survey, although in both cases the statistical power to detect a difference was clearly low due to sample size (Benoît, 2006; Bourdages et al., 2007). However, within the survey overlap area, catch numbers in the annual bottom-trawl surveys have increased since the early 2000s, coinciding with the deployment of the new vessel, *Teleost*, in both regions (Figure 6). The elevated catch levels supported a simple overlap analysis for the relative catch efficiency between *Teleost*-WIIA (used in the SGSL) and *Teleost*-Campelen (used in the NGSL).

The overlap analysis generally requires a sufficient amount of samples to fit a statistical distribution of the expected catch within each substratum, year, and length bin for the estimation of underlying density. The analysis was applied to 2006-2020 as the annual surveys in 2004 and 2005 featured reduced number of tows in the overlap area in part due to the comparative fishing experiments. The two surveys resulted in similar spatial distribution patterns, as illustrated for four length groups: 0-50cm, 51-70cm, 71-90cm and 90+cm, although catches were small overall and there was a visible scaling difference wherein the WIIA had larger catches for all lengths (Figure 7). To improve model efficiency given limited data, catch numbers were aggregated into 5cm length bins and the model was fit to the length range of 30-110cm, excluding mostly zeros for extremely large and small sizes and resulting in 17 length bins in total. The catch-length compositions from the two surveys in the same area were comparable, indicating some similarity in length selectivity of the two gears, although the SGSL survey had relatively higher catch efficiency (Figure 8). Average catch per tow showed a similar scale and annual trend for large halibut but for smaller sizes, the SGSL survey tended to have more sporadic high catches (Figures 6 and 9), especially in 2011, 2016 and 2020.

The overlap analysis fits a catch distribution for each substratum, year and length bin for the estimation of a shared population density within the group. The division of three substrata increased the resolution of the estimated population density over the overlap area, while the aggregation of catch over length both improved data quality and reduced model complexity. The conversion factor was developed to calibrate *Teleost*-WIIA to *Teleost*-Campelen (as the standard gear) and the offset term included tows standardization terms prior to modeling such as the proportion of subsampling and the swept area in km^2. For both candidate models applicable to the overlap analysis, BI0 and BI2, combinations of different settings of the over-dispersion and zero-inflation parameters were tested. Model diagnostics guided a determination of a common over-dispersion parameter for a negative binomial distribution (i.e., without zero-inflation). This is also reasonable as highly parameterized models tend to over-fit given small sample sizes.

Both models were fit to the data for a selection of length-dependency of the conversion factor. Results were nearly identical in terms of AIC (Table 8), with BI0 having slight advantage resulting from a simpler form. Estimations from both models were within the confidence intervals of each other and BI2 with a length smoothing effect also resulted in more estimation uncertainty as the length effect was not important enough to improve model fit (Figure 10). *Teleost*-WIIA was about 2.4 times more efficient than Teleost-Campelen for all lengths (estimated conversion by BI0 was 2.438 with a standard error of 0.272).

Model diagnostics undertaken using *DHARMa* indicated a suitable fit for the negative binomial distribution assumption. In addition, simulated catches using parameters estimated by the fitted model were assessed against observed catches to further test over-dispersion and zero-inflation. The residual diagnostics for the overlap analysis of the Atlantic halibut passed all tests (Figure 11).

Despite a suitable model fit (Figure 11) and a reasonable estimated conversion (Figure 10), the validity of the analysis rests on the assumption that the NGSL and SGSL surveys sampled the same underlying population density. This assumption would be violated if there was migration into or out of the overlap area during the one-month interval (average) between when the area is sampled by the SGSL and NGSL surveys. This could create a bias in the estimation, and more importantly, refute the possibility of combining the two surveys into one population index within GSL if the movements result in a non-negligible amount of fish effectively being double counted or unsampled due to the movements. Spatial distribution patterns of the resulting catches from the two surveys for Atlantic halibut were not extremely different, however previous analyses for Greenland Halibut have shown important systematic differences in density in the overlap area between the two surveys not due to differences in gear (Yin and Benoît, 2022).

3.2 Redfish: Comparative Fishing Analysis

The comparative fishing analysis was applied to each of the five comparative fishing experiments separately to generate pairwise relative catch efficiencies for redfish, *Sebastes spp*. The analyses presented here do not distinguish between *Sebastes mentella* and *Sebastes fasciatus* given that this species split is done post hoc based on additional sampling that was not available for the SGSL survey. Accounting for the species split in analyses of the NGSL comparative data would also not have been straightforward given uncertainties and biases inherent in estimating the proportion of each species in catches (Senay et al., 2021). The analyses used effective pairs in which at least one redfish was caught by either vessel-gear within the pair. The two comparative fishing experiments in the NGSL resulted in good amounts of effective data (Figures 12, 13), while positive catches from the three SGSL surveys were almost exclusively found in the Laurentian Channel in the overlap area (Figures 14, 15, 16), resulting in relatively few effective pairs (Table 9). The ratio estimator (ratio between total catches of the old and new vessel-gears regardless of length) indicated relatively higher catch efficiency for new vessel-gears in general, except in the SGSL 1992 survey where Lady Hammond-WIIA resulted in higher total catches than *Alfred Needler*-WIIA, and in the SGSL 2004-2005 survey where there was essentially no difference (Table 9).

For each of the five comparative fishing experiment analyses, catches were binned every 1cm within a length range selected based on the minimum and maximum measurements, the candidate binomial and beta-binomial models were then fit to the data, and the best

model was selected among all properly converged models to give the lowest AIC. Table 10 presents the difference in AIC from the best model for each of the 13 candidate models for each of the five analyses. The data did not support successful estimation of more complicated models such as BI4, BB6 and BB7; although some of these models returned results, the hessian matrices were not positive definite, possibly due to excessive parameterization or unsuitable overdispersion structure (in comparison with less complicated models). In all five analyses, the beta-binomial models generally performed better than their binomial counterparts, and the most complicated model among all converged models were favored, indicating both significant length effects and station effects.

For each analysis, the estimated proportion of catch by the old vessel, $p_{Ai}(l)$, was compared among all converged models and against the sample proportion of catch over length (computed for each station and for an average across stations) in order to assess different model assumptions (Figures 17-21). The proportions of catch over length were inconsistent among stations as exemplified by the spread of station-wise sample proportions (small dots in the figures). They were also non-monotonic along length, with sample averages showing vastly different trends at different length intervals (small circles in the figures), e.g., in NGSL 2004-2005 between length groups 15-30cm and 30-45cm (Figure 18), and in SGSL 1985 between 5-15cm and 15-30cm groups (Figure 19). This could be due to different interactions with the gears between different age groups of redfish.

The AIC-based model selection did not always favor the smoothest estimation. In the NGSL 1990 survey, Lady Hammond-WIIA was slightly more efficient than *Alfred Needler*-URI for redfish greater than 20cm, but for smaller sizes, *Alfred Needler*-URI was much more efficient (Figure 17). Notably, the simple ratio estimator ignoring the length effect was heavily biased by the smaller size range due to inflated catch numbers of small fish (Table 10). For redfish under 20cm, relative catch efficiency increases with length and this length effect plateaued between 20-40cm. Models with a smoothing length effect indicated this similar trend despite small variations in smoothness. However, the length effect was not significant globally. In NGSL 2004-2005, *Teleost*-Campelen was consistently more efficient (Figure 18). The three SGSL surveys had small data amounts and estimation quality was less ideal. Estimation from the analyses indicated that EE Prince-Yankee was less efficient at catching small fish (<15cm) than Lady Hammond-WIIA (Figure 19). The difference in catchability was not as evident for larger sizes (>15cm) based on the sample proportions of catch, but the model estimations projected some length effect nevertheless; the model performance was likely impacted by a cluster of pairs within 27-30cm where Lady Hammond-WIIA consistently caught more redfish. Lady Hammond-WIIA was more efficient for the 20-40cm range compared to *Alfred Needler*-WIIA, but the relative catch efficiency decreased for smaller and larger sizes (Figure 20). The comparative fishing experiment between *Alfred Needler*-WIIA and *Teleost*-WIIA indicated a similar catchability between gears, as effective pairs did not provide strong evidence of disparity from the sample proportions (Figure 21). The estimated relative catch efficiency (within the 5-45cm length range) between the vessel-gear pairs in the five comparative fishing analyses were derived from the estimated catch proportions and their standard deviations were calculated via a Delta method implemented in TMB (Figure 22). Estimates for the three southern comparative fishing experiments generally had higher levels of uncertainty due to relatively smaller sample sizes (number of effective pairs).

Previous analyses of the comparative fishing data for the SGSL had concluded there were no differences in efficiency, length-dependent or length-aggregated, among any of the vessel-gear tandems (Benoît and Swain, 2003; Benoît, 2006). This contrasts with the length-dependent patterns estimated here. For the NGSL, where redfish were much more frequently encountered during comparative fishing, the results obtained here for the *Alfred Needler-Teleost* experiments (Fig. 21) are very similar with those obtained by Bourdages et al. (2007), who concluded that the *Teleost* was about 2.29 times more efficient, regardless of length.

Prediction residuals were derived from the best models in each comparative fishing analysis and the normalized randomized quantile residuals were calculated and subsequently evaluated against stations and lengths for detecting potential estimation bias due to model misfit. For the two northern surveys, these residual diagnostics did not show extreme deviations or indications of model misspecification (Figures 23, 24). Residuals aggregated by station were generally more erratic than by lengths, showing some under-representation of the station-level differences despite a statistical random effect explicitly modeling station variations. Diagnostics for the three southern comparative fishing analyses, characterized by fewer effective stations, indicated elevated bias levels for some length bins but with no systematic patterns and mostly within reasonable ranges (Figures 25, 26, 27). Overall, model fits were acceptable.

Potential improvements could be made to better capture the station-level differences and address the connection between lengths. For example, redfish can occur in large three-dimensional aggregations off bottom. If the vertical extent or density of these aggregations vary systematically, for example with depth or with local abundance, then differences in the vertical opening of the trawls could result in in relative catchability that covaries with these factors. The inconsistent length effect as observed in the paired catches could also be caused by different spatial distributions of different size (or age) classes that were not adequately sampled. In such cases, further smoothing may be required, either by restricting the smoothness parameter, or by reducing the number of smoothing knots for a coarser smoothing grid to bound this parameter. However, it is not clear how to determine whether more constrained smoothing is appropriate and to which degree. Additional constraints could be placed on the shape of the relative catch efficiency over length, such as using a monotonic function, given appropriate supporting biological and mechanistic evidence, or even fitting to parametric functions (Bourdages et al., 2007), but any additional model assumptions require further study for justification.

3.3 Redfish: Integrated Model

The annual RV surveys in the SGSL and NGSL mostly overlap in the Laurentian Channel (Figure 3), where redfish catches can support the overlap analysis. Catches in the surveys were intermediate from 1984 to the early 1990s, considerably lower for most of the period from 1995 to 2012 and have since increased considerably (Figure 28). Historical surveys including the five comparative fishing experiments have been sampled by six vessel-gear combinations during 1984-2020, and despite potential differences in relative catch efficiency, catch trends within the overlap area (Figure 29) and annual spatial distribution patterns (Figures 30, 31, 32) were both similar for different length groups, suggesting there is consistency in the relative catch efficiencies among gears over time. In order to quantify the relative catch efficiencies among all six vessel-gear combinations, the integrated model combining the overlap analysis and the above comparative fishing analyses was applied to

the catches within the overlap area and catches from both SGSL and NGSL comparative fishing experiments.

Given the apparent general suitability of the comparative fishing experiments to estimate the relative catchability functions, we did not undertake an estimation of those functions based on an overlap analysis alone, as was done for Atlantic halibut. Instead, we went directly to the integrated analyses as these had the potential to improve the estimates. In the integrated analysis for redfish, the same three substrata partitioning the overlap area were used (Figure 5). Catches were aggregated in 2cm length bins and the modeling length range was restricted within 5-46cm. This limited the total number of fixed parameters and hence greatly reduced computation to a reasonable degree. For the conversion, all vessel-gears were calibrated to *Teleost*-Campelen equivalent. Cubic smoothing over length was based on 10 knots (K = 9) even if the length range for this dataset was more restricted than for Atlantic halibut. While various settings were tested, model diagnostics favored a uniform over-dispersion parameter among all substrata, years and length bins, and a set of zero-inflation parameters that varied according to each year and length.

Estimated catch efficiency for each gear relative to the standard, *Teleost*-Campelen, is presented in Figure 33 on the log scale. *Teleost*-Campelen is the most efficient among all gears, especially for small sizes. *Teleost*-WIIA, in comparison, catches around 1/7 for all lengths; relative catch efficiencies for *Alfred Needler*-URI and *Alfred Needler*-WIIA featured a slight increase after 30cm and 20cm, respectively, but approximately constant for smaller sizes; within the modelled length range, catch efficiencies for Lady Hammond-WIIA and EE Prince-Yankee were estimated to increase monotonically with length.

In order to compare with results from separate comparative fishing analyses, the estimated conversions to *Teleost*-Campelen from the integrated model were transformed into pairwise calibrations corresponding to the five pair of gears in the comparative fishing experiments, and their estimation uncertainty was calculated via Delta method (Figure 34). Conversion factors estimated by the integrated analysis were mostly smoother than the comparative fishing analyses, especially for the three vessel-gears pairs in the SGSL survey where the three comparative fishing experiments did not result in sufficient amounts of data. The integrated analysis and the separate comparative fishing analyses both estimated conversion factors with similar general trends for length dependency, except for the case of *Alfred Needler*-WIIA vs. *Teleost*-WIIA, where the two methods indicated different catchability trends for fish above 30cm. The scale of conversions (to *Teleost*-Campelen) for the two gears used in the NGSL remained similar to results from comparative fishing analyses, but for the three gears used in the SGSL, scaling differences were significant in some cases, especially in the case of *Teleost*-WIIA. This may result from an inconsistent performance of the same gear, Lady Hammond-WIIA, between its deployments in the two different surveys (NSGL and SGSL). However this seems unlikely given that the vessel was a charter boat that very likely operated identically in the two surveys. An alternative possibility is a change of underlying density within the overlap area between the two surveys due to fish migration in the intervening time between the two surveys. Both cases would consequently contribute to a bias when converting the gears used in the SGSL, especially via the multiplication method. Conforming the NGSL and SGSL Lady Hammond-WIIA catches resulted in the gaps in the cases of Lady Hammond-WIIA vs. *Alfred Needler*-URI and Lady Hammond-WIIA vs. *Alfred Needler*-WIIA.

Quantile residuals were computed for the comparative fishing pairs and checked against length bins to detect extreme or patterned deviations (Figures 35 and 36). The randomized normalized quantile residuals were calculated for a single tow within each pair to avoid a symmetric duplication. Residuals did not indicate severe problems for the beta-binomial fit of the paired catches.

For the fit of negative binomial catches, the more rigorous *DHARMa* residual diagnostics were used (Figure 37). Goodness-of-fit tests on the scaled residuals demonstrated an adequate fit of the zero rate and variance in the original data, but also indicated a deviation from the overall distribution assumption (p-value for the KS test in Figure 37). This was a persistent issue among a varying degree of model assumptions we tested and possibly due to extremely dispersed catch numbers (catches of small sizes of redfish were highly inflated, in particular). To better capture the catch dispersion, the model may require a distribution with a heavier tail than the negative binomial distribution. However, this deviation did not seem to impact the estimation of conversion functions when comparing multiple models, and a comparison of predicted and observed catches for each length bin did not indicate significant bias, either (Figure 38). The zero-inflation (reverse of presence probability) within the overlap area were highest for small and large redfish (Figure 39), and the annual trend (Figures 39 and 40) agreed with yearly catches (Figure 29).

Both the separate comparative fishing analyses and the integrated analysis present reasonable model diagnostics and appear appropriate. Choosing between the two to establish the conversion functions to use to standardize the surveys is not straightforward. On one hand, the integrated analysis could be considered superior because it incorporates more data and more information on relative catchabilities. Furthermore, this additional information was cumulated over time and may therefore be less subject to possible biases inherent in the data from the comparative fishing experiments, each of which was effectively unreplicated. (In the absence of replication, it is not possible to conclude that if the comparative fishing experiments were repeated one of more times, they would produce the same result.). However, on the other hand, the validity of the integrated models rests on the assumption that SGSL and NGSL surveys fished the same densities of redfish in the substrata of the overlap area in a given year. Changes in density during the four weeks that typically separate the surveys would incorrectly be interpreted by the model as a difference in catchability, while the limited number of survey stations within this area and resulting catches can hardly be used to validate, let alone quantify, any potential changes. In the absence of a firm basis to choose the results of one method over the other, we applied the conversions functions estimated by each to the survey catches to estimate standardized survey indices and to examine the consequences of the choice of model. The differences in surveys indices were small and not systematic (Figure 41). Choosing one set of conversion functions over the other would not change the interpretation of the indices, whether catches of all sizes are aggregate together or aggregated by the size groups typically used in the assessment for the stock. Furthermore, the difference in annual relative length frequencies was almost imperceptible (Figure 42). These results clearly indicate that the choice of conversion functions from among the two analyses is inconsequential. This result likely stems from the fact that the difference in conversion functions were most pronounced for the SGSL, which has a minor contribution to the indices compared to the NGSL survey, and for the less abundant largest sizes in the case of the SGSL *Needler* WIIA-*Teleost* WIIA comparison.

4 CONCLUSIONS

Paired-tow comparative fishing experiments are the gold-standard for calibrating for changes in survey vessel, gear or procedure because the design accounts for the influence of many factors that can affect catches. However the experiments are almost never replicated, raising the potential for incorrect estimates of relative catch-efficiencies. Furthermore, for species that were rare in past comparative fishing experiments but have since become more abundant, it is often not possible to accurately standardize time series based on comparative fishing alone. Analyses based on survey spatiotemporal overlap, and especially ones integrating comparative fishing results have the potential to attenuate or solve these issues. For instance, we were able to intercalibrate SGSL and NGSL surveys for Atlantic halibut. A reanalysis of comparative fishing data for redfish with more powerful statistical methods allowed for the identification of capture-efficiency effects in the SGSL that had not previously been identified. Importantly, a comparison of result from comparative fishing alone and from the integrated model identified some non-negligible discrepancies that warrant further investigation. Fortunately these do not seem impactful for the survey indices.

The application of the spatiotemporal model to GSL survey data, and in particular the development of the integrated model are key novel contributions of our work. These methods should be broadly applicable to the calibration of other surveys that overlap spatially and that can be assumed to sample the same underlying densities of fish.

5 ACKNOWLEDGEMENTS

Caroline Senay and Daniel Ricard provided detailed and thoughtful reviews of the penultimate draft of this report. We also thank the handling editor, Jaclyn Hill.

6 REFERENCES

Bagley, N.W.; Horn, P.L.; Hurst, R.J.; Jones, E.; Parker, S.J.; Starr, P.J. 2015. A review of current international approaches to standardisation and calibration in trawl survey time series. New Zealand Fisheries Assessment Report 2015/46. 54 p.

Benoît, H.P., and Swain, D.P. 2003. 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. 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., Ouellette-Plante, J., Yin, Y., and Brassard, C. (in review). Review of the assessment framework for Atlantic cod in NAFO 3Pn4RS: fishery independent surveys. DFO Can. Sci. Advis. Sec. Res. Doc. 2021/XXX. iv + XX p.

Bourdages, H., Savard, L., Archambault, D., and Valois, S. 2007. Results from the August 2004 and 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.

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 Bulletin 6: 25–42.

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

Fowler, G.M. and Showell, M.A. 2009. Calibration of bottom trawl survey vessels: comparative fishing between the Alfred Needler and Teleost on the Scotian Shelf during the summer of 2005. Can. Tech. Rep. Fish. Aquat. Sci. 2824: iv + 25 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).

Hartig, F. 2020. DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. R package version 0.3.3.0. https://CRAN.R-project.org/package=DHARMa.

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.

Marsaglia, G., Tsang, W.W. and Wang, J. 2003. Evaluating Kolmogorov's Distribution. J. Stat. Softw. 8: 1–4. DOI:https://doi.org/10.18637/jss.v008.i18.

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.

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

Savoie L. 2014. Results from the 2012 and 2013 sentinel bottom-trawl surveys in the southern Gulf of St. Lawrence and comparisons with previous 2003 to 2011 surveys. DFO Can. Sci. Advis. Sec. Res. Doc. 2014/054: v + 63 p.

Senay, C., Bermingham, T., Parent, G.J., Benoît, H. P., Parent, E., Bourret, A. 2021. Identifying two Redfish species, *Sebastes mentella* and *S. fasciatus*, in fishery and survey catches using anal fin ray count in Units 1 and 2. Can. Tech. Rep. Fish. Aquat. Sci. 3445.

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 Ward, E.J. 2014. Accounting for vessel effects when standardizing catch rates from cooperative surveys. Fish. Res. 155: 168-176.

Thorson, J.T., and Minto, C. 2015. Mixed effects: a unifying framework for statistical modelling in fisheries biology. ICES Journal of Marine Science, 72(5), pp.1245-1256.

Warren, W.G. 1997. Report on the comparative fishing trial between the Gadus Atlantica and Teleost. NAFO Scientific council studies, 29: 81-92.

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.

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. vii + 45 p.

7 TABLES AND FIGURES

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). Note that exceptionally in 2003, the CCGS Wilfred Templemen was used in the SGSL survey. This vessel is the sister ship to the CCGS Alfred Needler, because these vessels shared the same parameters, the W. Templeman is not included in the table.

	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		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	3.7-4.6
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: Pairs of vessel-gears and number of stations for the two northern Gulf of St. Lawrence comparative fishing experiments (i.e., NGSL 1990 and NGSL 2004-2005) and the three southern Gulf of St. Lawrence comparative fishing experiments (i.e., SGSL 1985, SGSL 1992, and SGSL 2004-2005).

CF Survey	**Old vessel-gear**	**New vessel-gear**	**Number of stations**
NGSL 1990	Lady Hammond WIIA	CCGS *Alfred Needler* URI	80
NGSL 2004-2005	CCGS *Alfred Needler* URI	CCGS *Teleost* Campelen	161
SGSL 1985	E.E. Prince Yankee	Lady Hammond WIIA	61
SGSL 1992	Lady Hammond WIIA	CCGS *Alfred Needler* WIIA	66
SGSL 2004-2005	CCGS *Alfred Needler* WIIA	CCGS *Teleost* WIIA	101

Table 4: 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	$log(\rho)$	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 5: 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	$log(\rho)$	$log(\phi)$	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	$\mathbf{X}_f^T\boldsymbol{\beta}_f + \mathbf{X}_r^T\mathbf{b}$	γ_0	smoothing/constant	not considered
BB_3	$\mathbf{X}_f^T\boldsymbol{\beta}_f + \mathbf{X}_r^T\mathbf{b}$	$\mathbf{X}_f^T\boldsymbol{\gamma} + \mathbf{X}_r^T\mathbf{g}$	smoothing/smoothing	not considered
BB_4	$\mathbf{X}_f^T\boldsymbol{\beta}_f + \mathbf{X}_r^T\mathbf{b} + \delta_{0,i}$	γ_0	smoothing/constant	intercept
BB_5	$\mathbf{X}_f^T\boldsymbol{\beta}_f + \mathbf{X}_r^T\mathbf{b} + \delta_{0,i}$	$\mathbf{X}_f^T\boldsymbol{\gamma} + \mathbf{X}_r^T\mathbf{g}$	smoothing/smoothing	intercept
BB_6	$\mathbf{X}_f^T(\boldsymbol{\beta}_f + \boldsymbol{\delta}_i) + \mathbf{X}_r^T(\mathbf{b} + \boldsymbol{\epsilon}_i)$	γ_0	smoothing/constant	intercept, smoother
BB_7	$\mathbf{X}_f^T(\boldsymbol{\beta}_f + \boldsymbol{\delta}_i) + \mathbf{X}_r^T(\mathbf{b} + \boldsymbol{\epsilon}_i)$	$\mathbf{X}_f^T\boldsymbol{\gamma} + \mathbf{X}_r^T\mathbf{g}$	smoothing/smoothing	intercept, smoother

Table 6: Number of annual survey tows within the overlap area for both northern and southern Gulf of St. Lawrence (NGSL and SGSL). The column headers identify the vessels and trawls. The vessels include E.E. Prince (EE), MV Lady Hammond (LH), CCGS Alfred Needler (AN) and CCGS Teleost (Te).

Year	**NGSL**			**SGSL**			
	LH|WIIA	**AN|URI**	**Te|Campelen**	**EE |Yankee**	**LH|WIIA**	**AN|WIIA**	**Te|WIIA**
1984	13			7			
1985	22			6	9		
1986	12				10		
1987	10				11		
1988	16				16		
1989	17				13		
1990	13	15			10		
1991		19			11		
1992		16			7	18	
1993		15				13	
1994		16				12	
1995		11				12	
1996		20				17	
1997		21				13	
1998		15				19	
1999		19				15	
2000		16				17	
2001		15				11	
2002		16				13	
2003		14				4	
2004		4	12			3	11
2005		11	15			9	7
2006			16				11
2007			15				14
2008			17				15
2009			15				12

2010	11	10
2011	14	8
2012	14	10
2013	13	7
2014	9	12
2015	16	13
2016	12	14
2017	10	12
2018	9	10
2019	11	12
2020	9	11

Table 7: Catches of Atlantic halibut from historical comparative fishing (CF) experiments. Number of effective stations is the number of stations where at least one halibut was caught by either gear within the pair. Number of total stations is the total number of successful comparative fishing stations during the survey. Average number of observations is the average number of effective pairs over the total number of 1cm length bins.

CF Experiment	# Effective Station	# Total Station	Average # Observations
NGSL 1990	6	80	0.0804
NGSL 2004-2005	33	161	0.6882
SGSL 1985	0	61	0
SGSL 1992	0	66	0
SGSL 2004-2005	11	101	0.2857

Table 8: Model selection for the overlap analysis of Atlantic halibut between a smoothing length effect and a conversion without length effect.

Model	Description	Negative log-likelihood	AIC Difference
BI0	Without length effect	1204.604	0
BI2	Smoothing length effect	1205.931	6.654

Table 9: Summary of redfish catch data from the five comparative fishing experiments. Effective stations are where a pair of tows resulted in at least one catch. The ratio estimator is the ratio between the total catches of the old and the new vessels (both effort-standardized).

CF Experiment	# Effective Stations	# Total Stations	Average # Observations	Ratio Estimator
NGSL 1990	79	80	35.7551	0.5954477
NGSL 2004-2005	139	161	35.5849	0.440715
SGSL 1985	18	61	6.255814	0.7261231
SGSL 1992	13	66	5.953488	1.363776
SGSL 2004-2005	25	101	5.023256	0.9844527

Table 10: Model selection for the five comparative fishing analyses: difference in AIC of all converged models from the best model (indicated by zeros in bold).

Model	**NGSL 1990**	**NGSL 2004-2005**	**SGSL 1985**	**SGSL 1992**	**SGSL 2004-2005**
BI0	12362	3618	518	310	274
BI1	3105	799	252	124	206
BI2	5432	3553	396	250	168
BI3	1199	661	144	48	102
BI4	-	-	-	-	-
BB0	1008	705	230	120	55
BB1	409	11	154	58	36
BB2	603	641	198	79	36
BB3	510	589	86	60	23
BB4	**0**	**0**	109	7	30
BB5	-	-	**0**	**0**	**0**
BB6	-	-	-	-	-
BB7	-	-	-	-	-

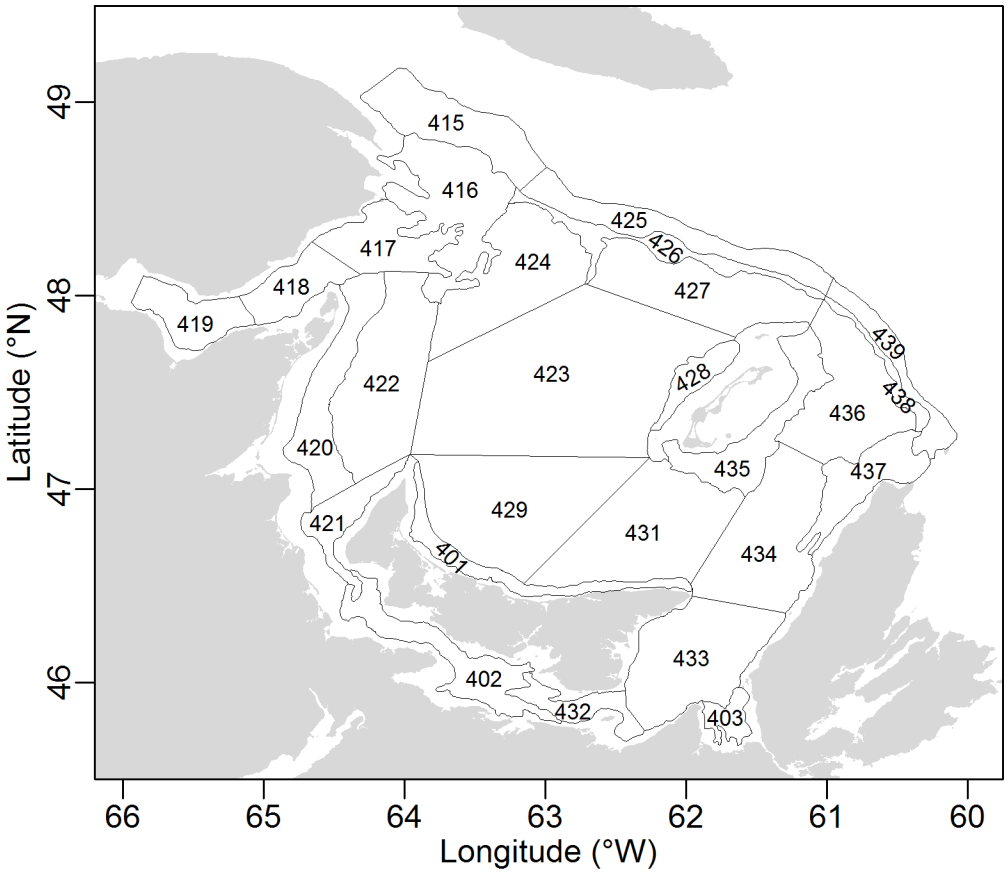

Figure 1: Stratification scheme for the southern Gulf of St. Lawrence multi-species bottom-trawl survey. The annual sampling design has included strata 415-439 since 1971 and strata 401-403 since 1984.

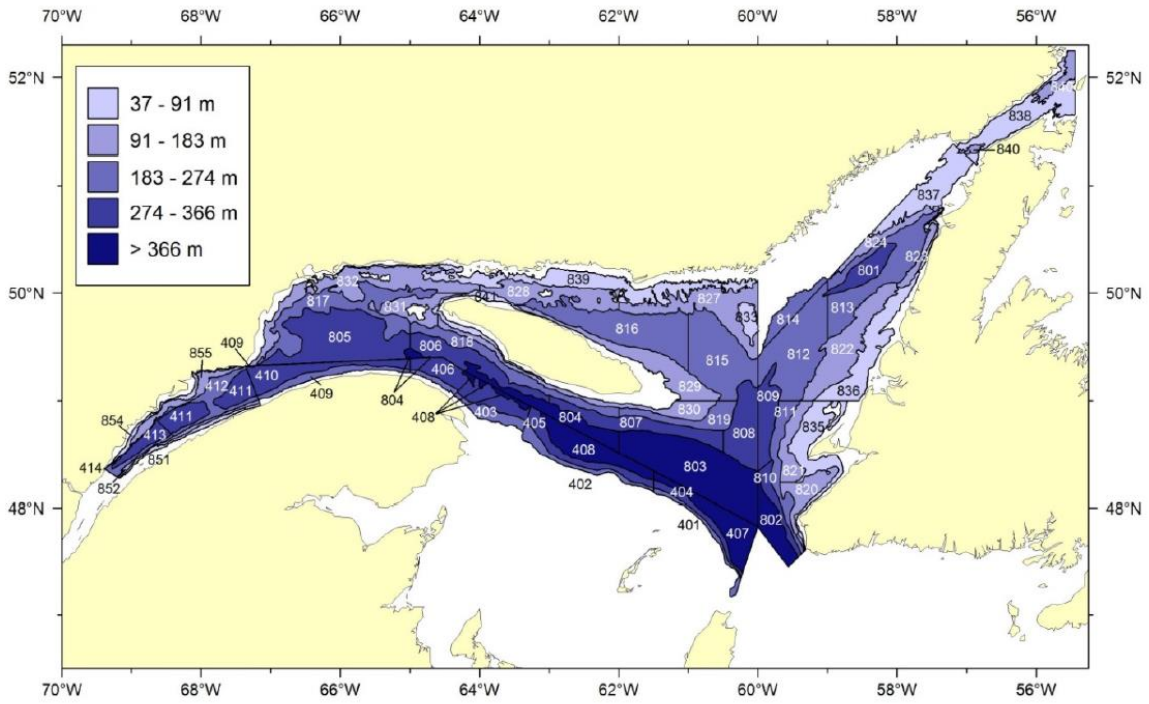

Figure 2: Stratification scheme for the northern Gulf of St. Lawrence multi-species bottom-trawl survey. The annual sampling design has included strata 401-408, 801-824 and 830 since 1984, and 827-829, 831 and 832 since 1985. Additional strata, located in NAFO area 3Ps (southwest Newfoundland) and sampled only in 1987 and 1993-2003 are not shown.

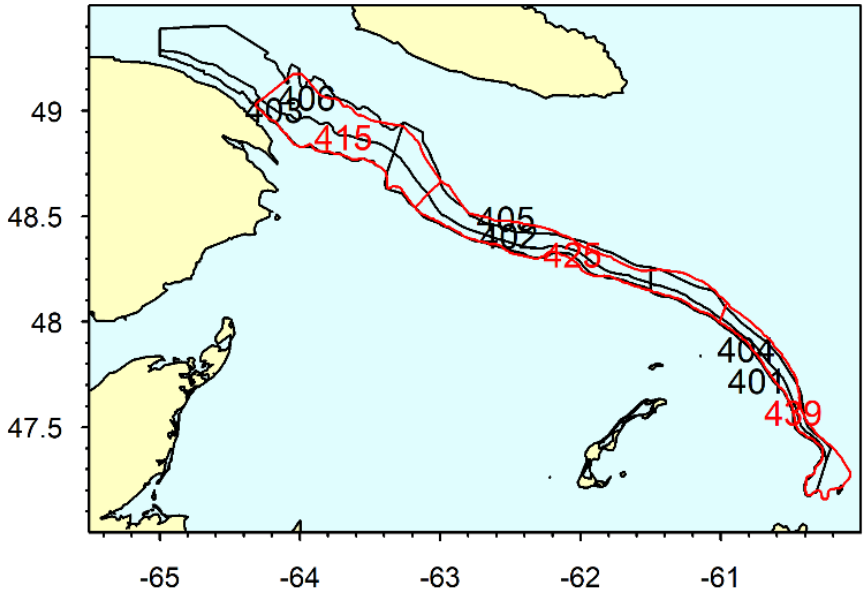

Figure 3: A close-up map of the strata from the southern Gulf of St. Lawrence (in red; strata 415, 425, 439) and northern Gulf of St. Lawrence (in black; strata 401-406) surveys in the area along the southern slope of the Laurentian channel where the two surveys overlap.

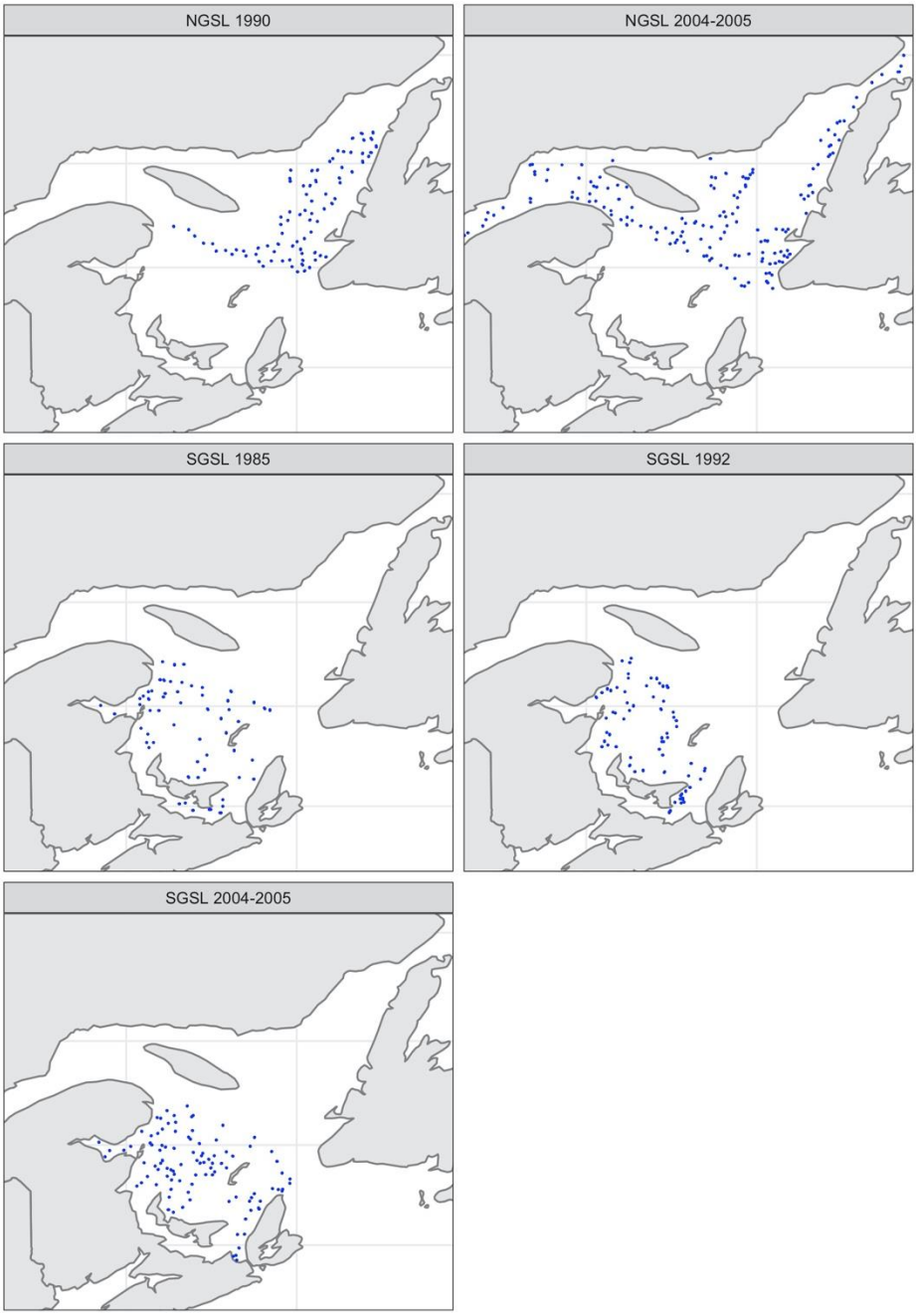

Figure 4: Survey tow locations for the five comparative fishing experiments, two in the northern Gulf of St. Lawrence (NGSL) and three in the southern Gulf of St. Lawrence (SGSL).

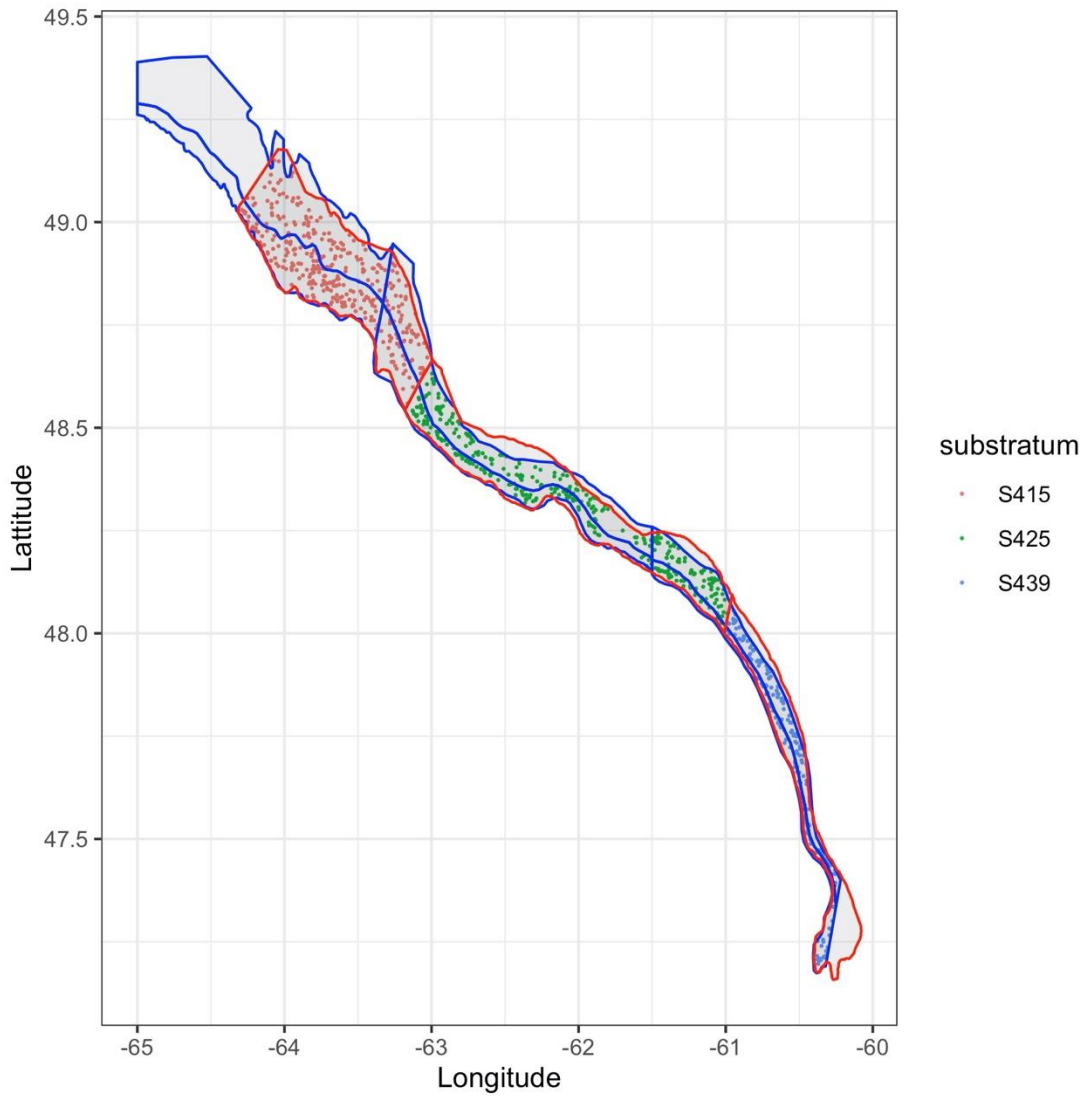

Figure 5: Tow locations within the overlap area from both the northern and southern Gulf of St. Lawrence (NGSL and SGSL) RV surveys and the scheme of their stratification into three substrata for the overlap analysis (colored points). Blue lines border the six NGSL RV survey strata within the area, red lines border the three SGSL RV survey strata; only survey tows within both surveys were included for the overlap analysis.

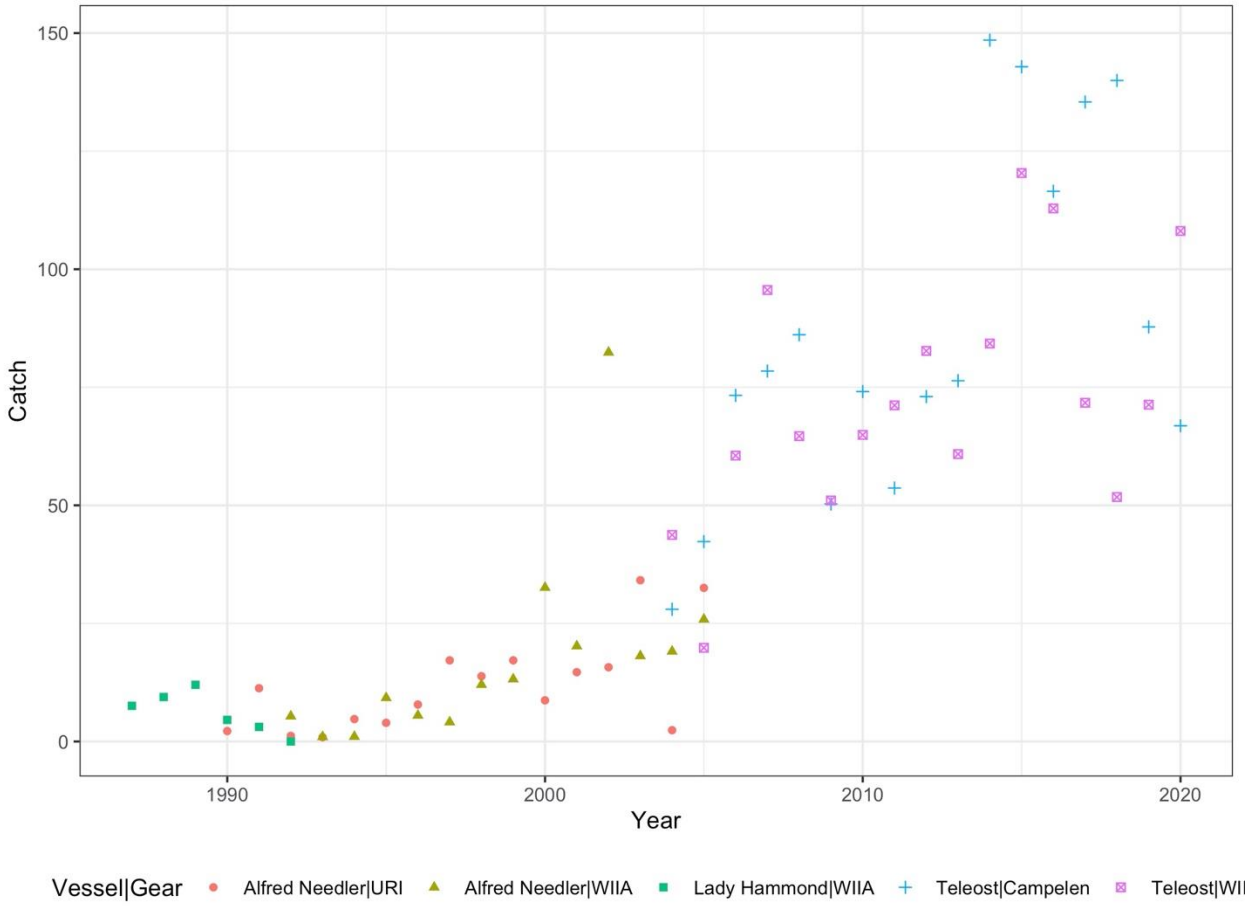

Figure 6: Total catches of Atlantic halibut by each vessel-gear from the annual RV surveys in both southern and northern Gulf of St. Lawrence (catch numbers were effort-adjusted according to the standard in the respective surveys).

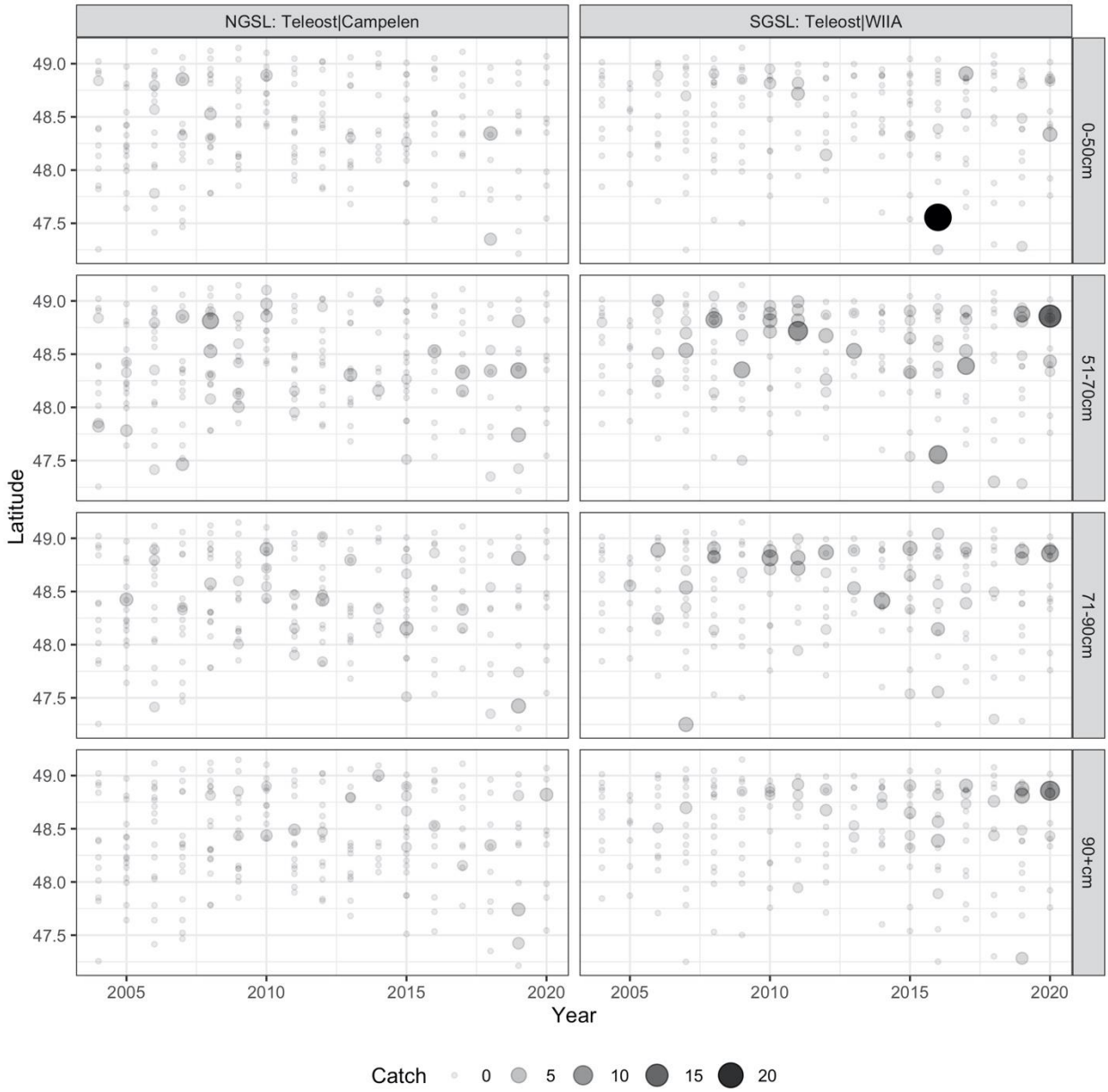

Figure 7: Latitude distribution of catches by *Teleost*-Campelen and *Teleost*-WIIA in the overlap area for four length groups of Atlantic halibut. Note that because of the shape and arrangement of the area, latitude acts as a proxy for location along the three substrata (see Figures 3 and 5).

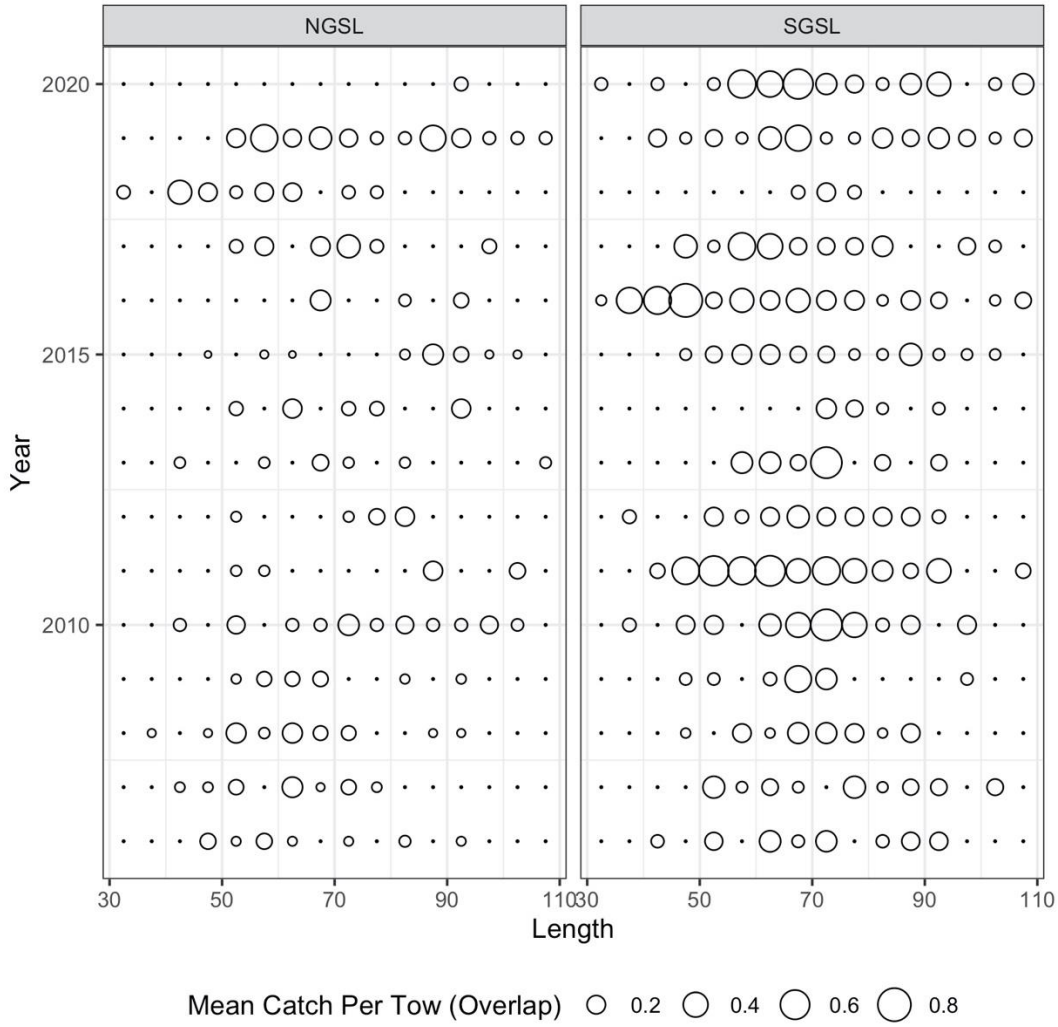

Figure 8: Catch-length composition for Atlantic halibut from the annual surveys in the northern and southern Gulf of St. Lawrence (NGSL and SGSL) in the overlap area. Catches were aggregated every 5cm between 30-110cm.

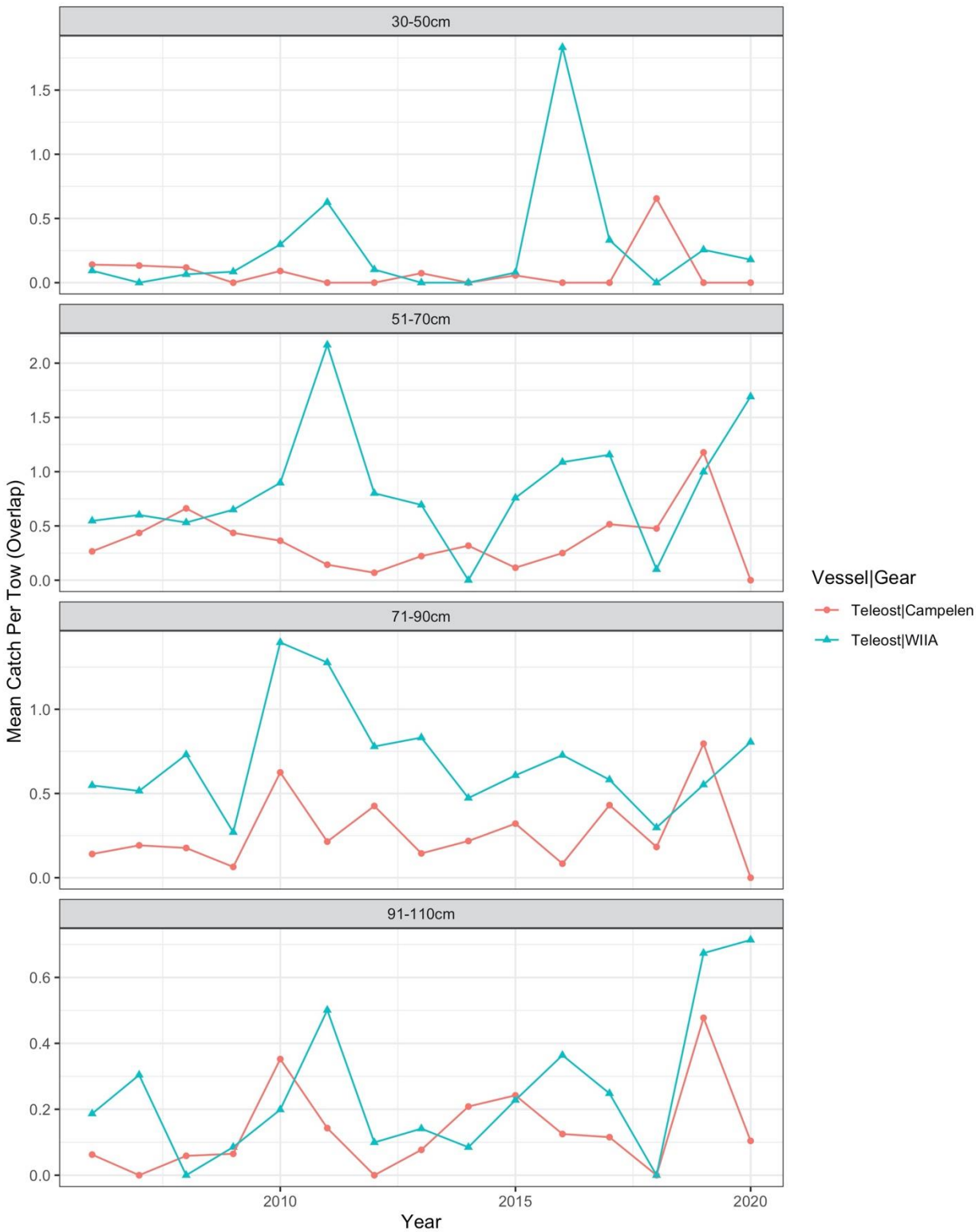

Figure 9: Annual trend of mean catch per tow for four length groups of Atlantic halibut using catches within the modeled range, within 2006 to 2020, and from the overlap area.

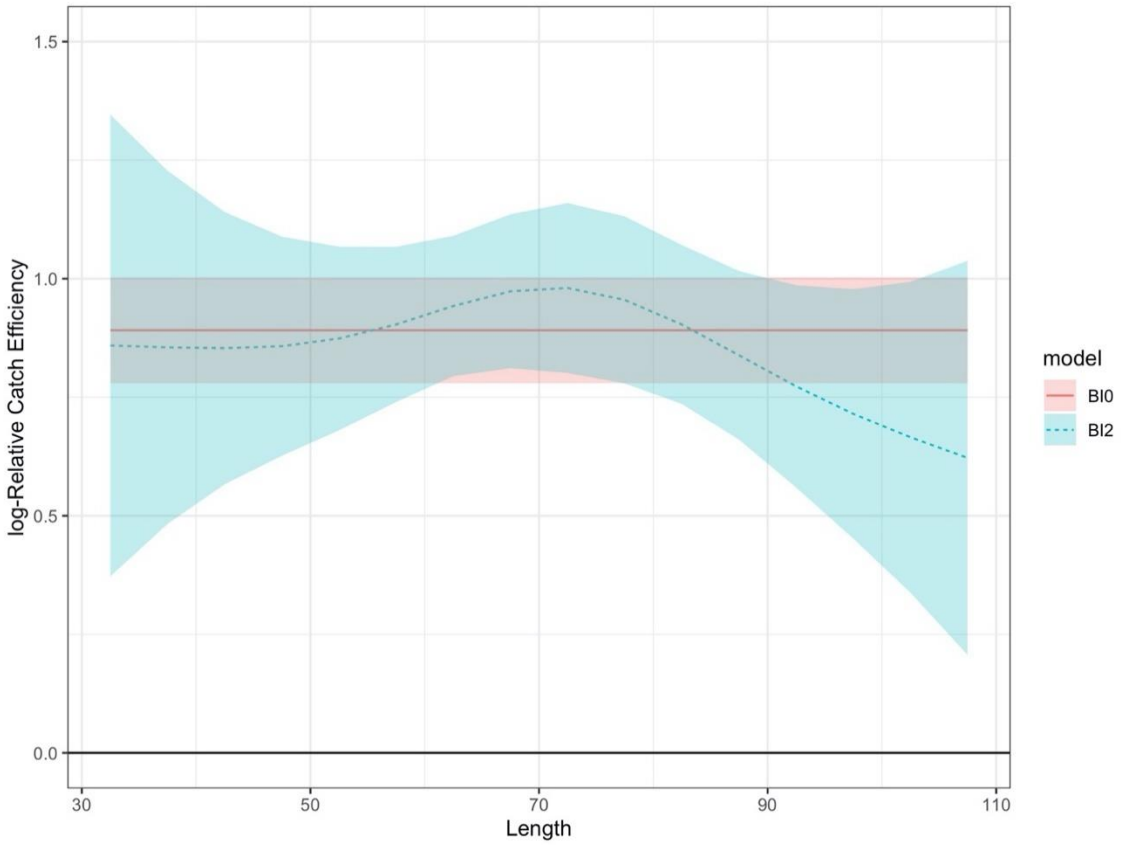

Figure 10: Estimated relative catch efficiencies between *Teleost*-WIIA and *Teleost*-Campelen on the log scale with one standard deviation by the two candidate models BI0 and BI2 in the overlap analysis of Atlantic halibut.

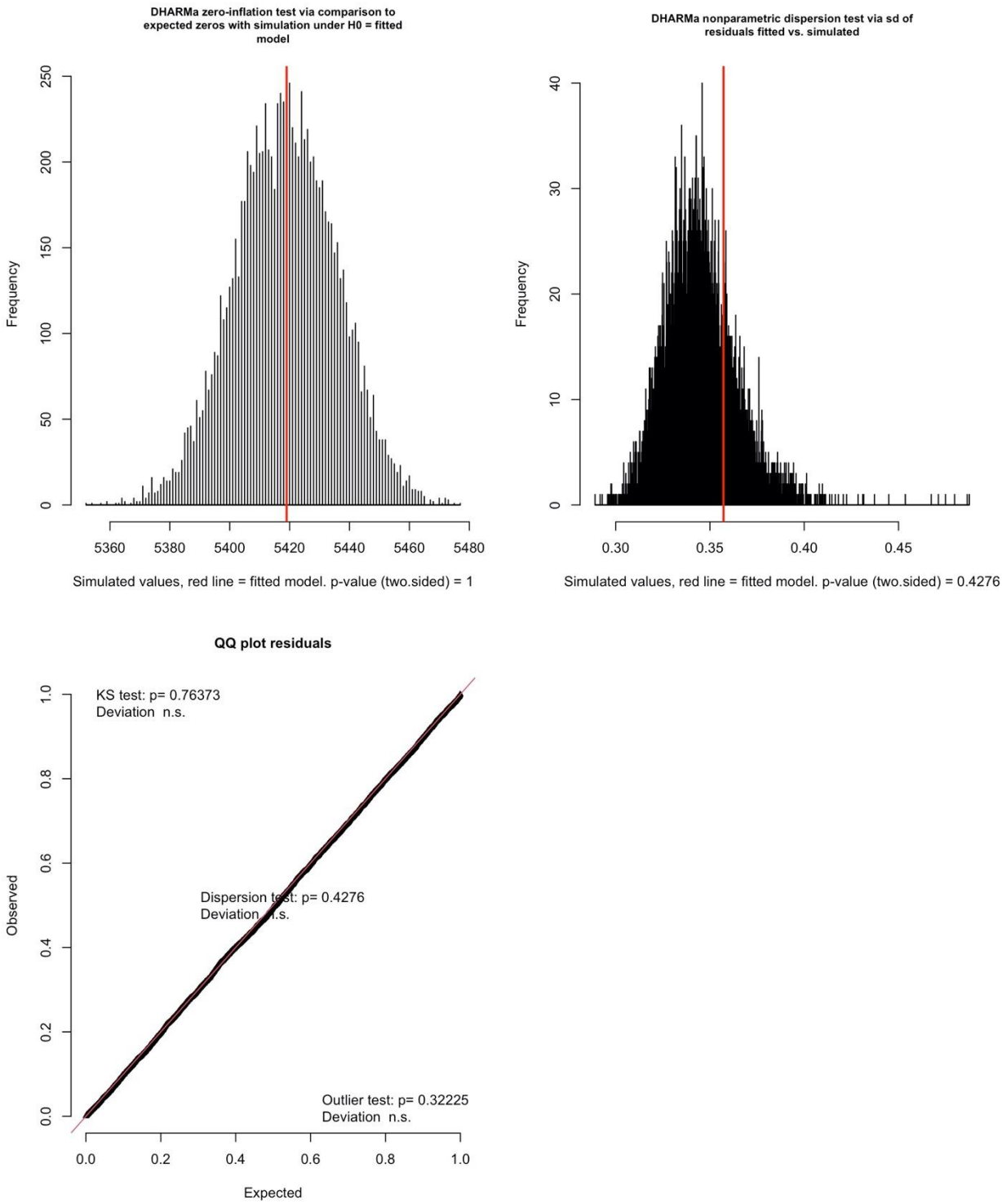

Figure 11: *DHARMa* residual diagnostics for the catch distribution assumption, including tests for zero-inflation (top-left panel), over-dispersion (top-right panel) and goodness-of-fit of the negative binomial distribution assumption (bottom-left panel), as outputted from the DHARMa package for R. The model past all tests, indicating a suitable fit.

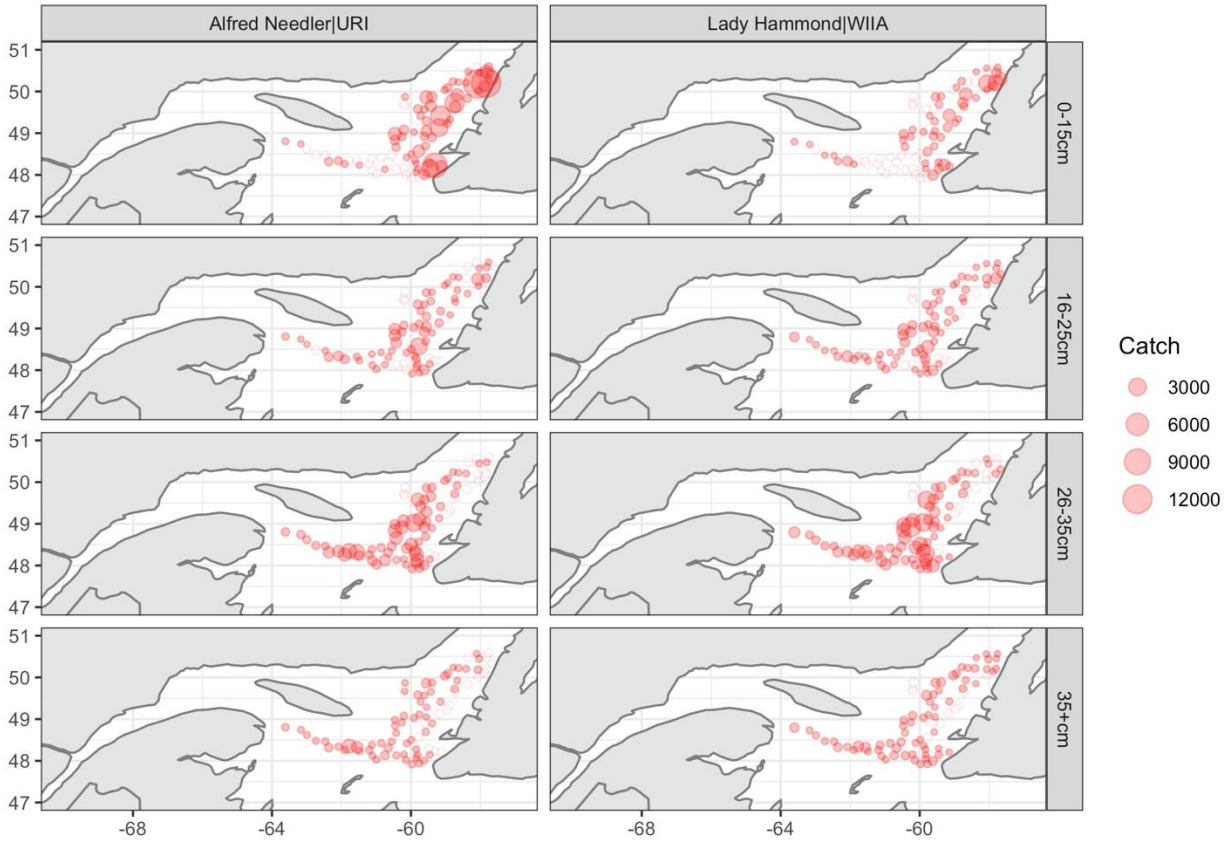

Figure 12: Catch of redfish (standardized number) by paired vessels from the comparative fishing experiment in the northern Gulf of St. Lawrence in 1990 in four length groups (solid red circles for positive catches and open circles for zero catches).

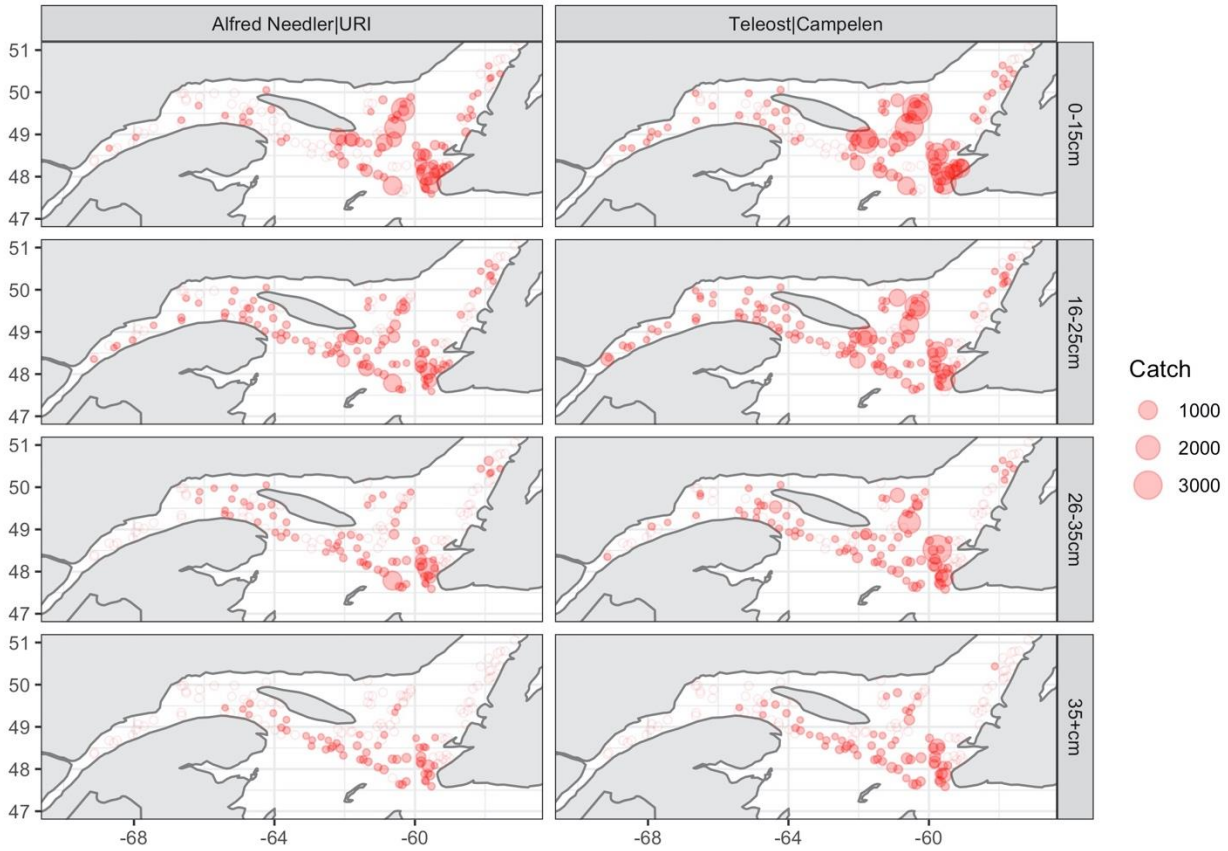

Figure 13: Catch of redfish (standardized number) by paired vessels from the comparative fishing experiment in the northern Gulf of St. Lawrence in 2004-2005 in four length groups (solid red circles for positive catches and open circles for zero catches).

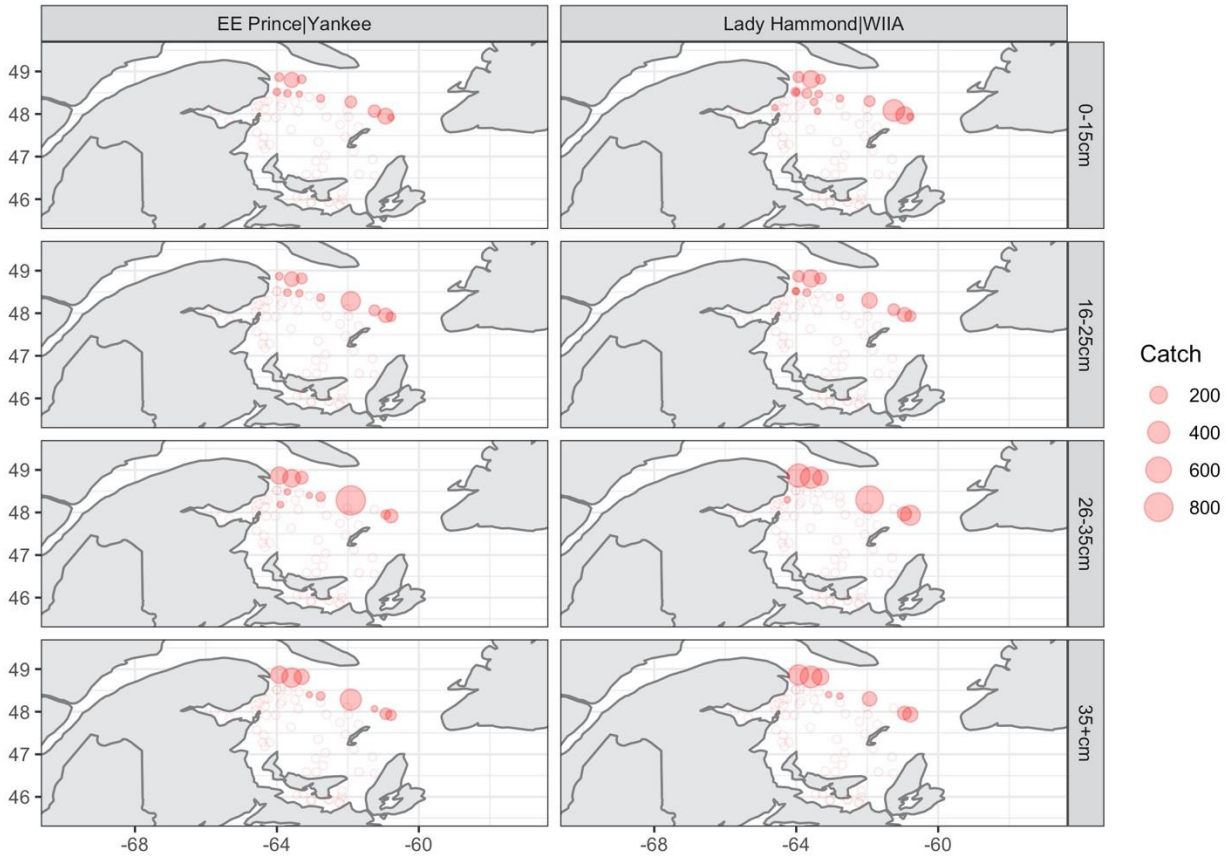

Figure 14: Catch of redfish (standardized number) by paired vessels from the comparative fishing experiment in the southern Gulf of St. Lawrence in 1985 in four length groups (solid red circles for positive catches and open circles for zero catches).

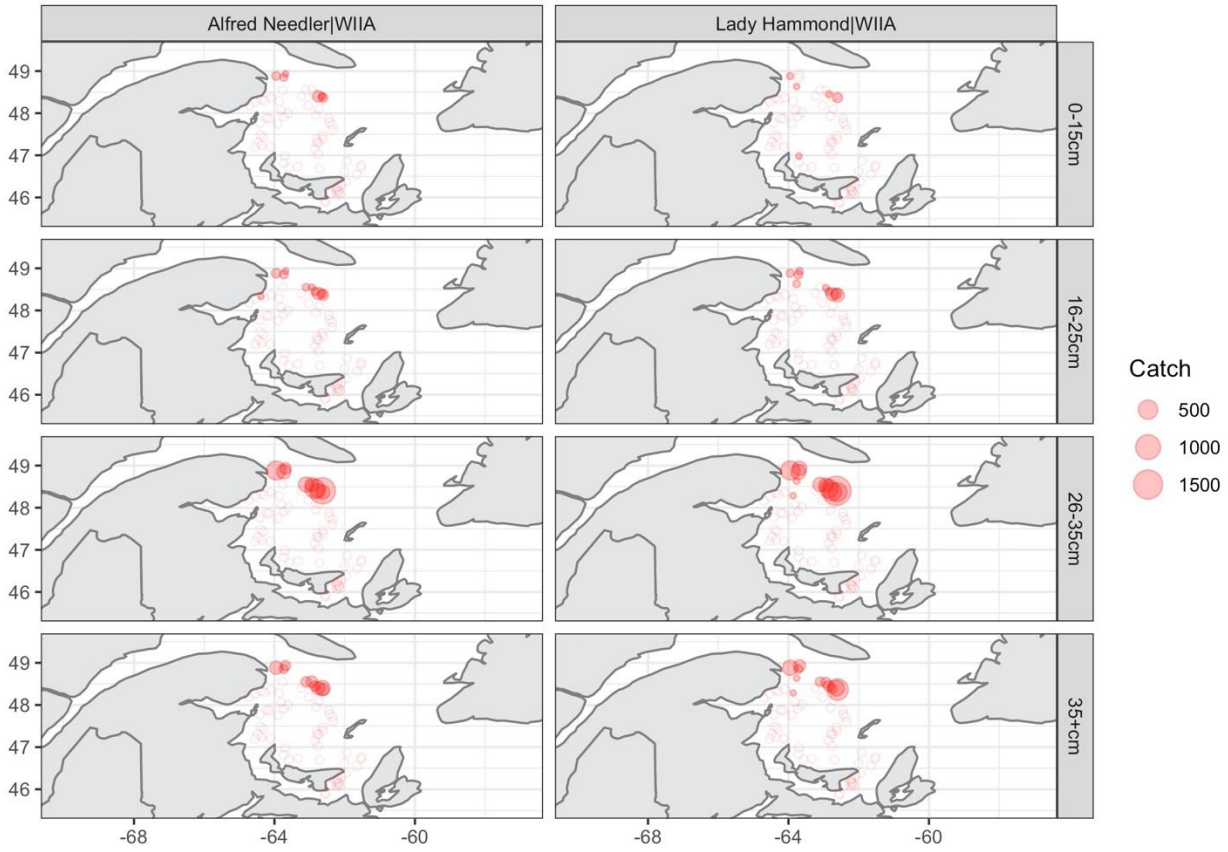

Figure 15: Catch of redfish (standardized number) by paired vessels from the comparative fishing experiment in the southern Gulf of St. Lawrence in 1992 in four length groups (solid red circles for positive catches and open circles for zero catches).

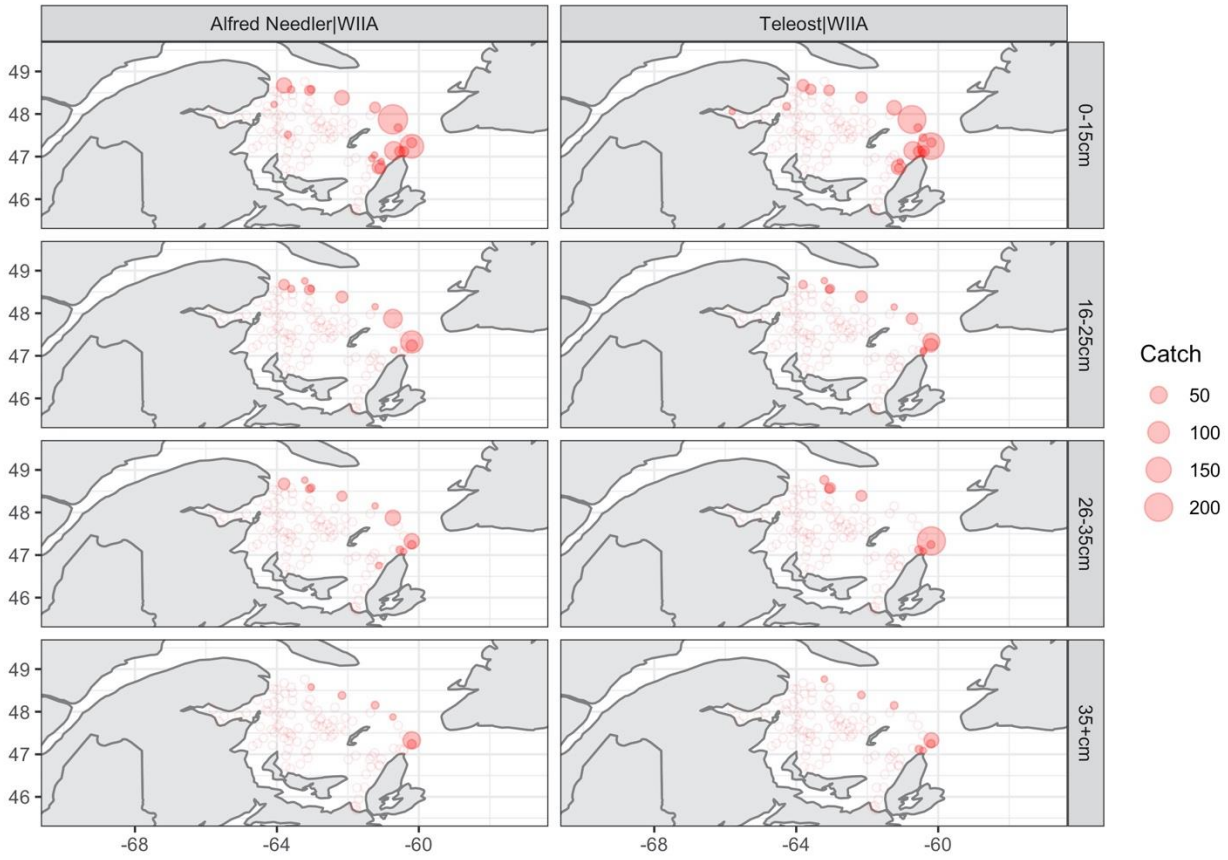

Figure 16: Catch of redfish (standardized number) by paired vessels from the comparative fishing experiment in the southern Gulf of St. Lawrence in 2004-2005 in four length groups (solid red circles for positive catches and open circles for zero catches).

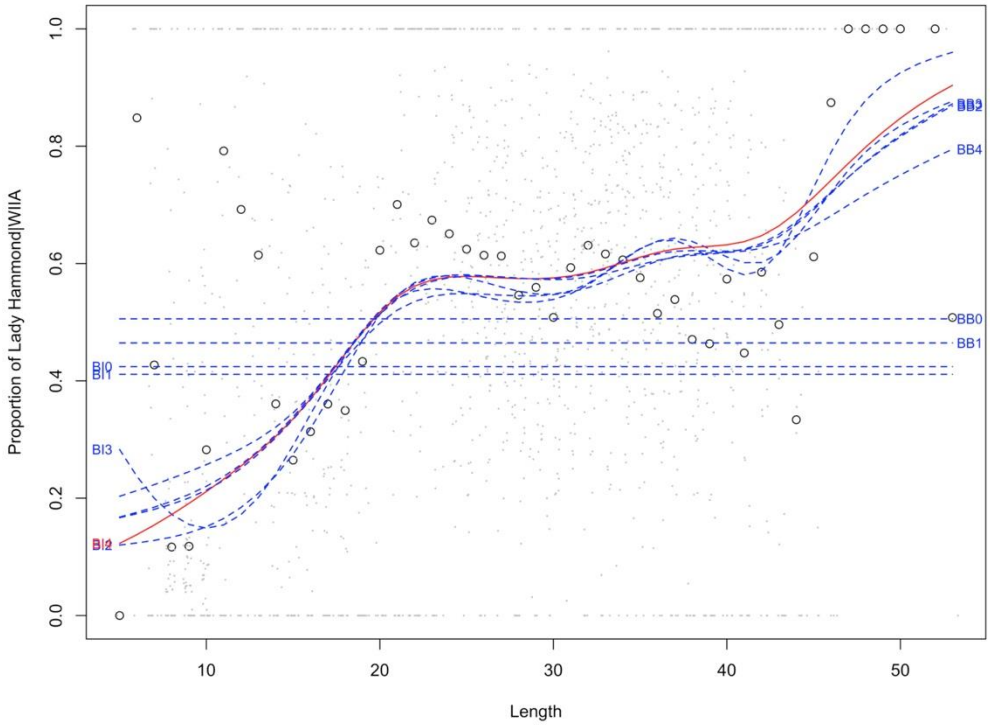

Figure 17: Comparative fishing analysis of NGSL 1990, between Lady Hammond-WIIA and *Alfred Needler*-URI: Estimated proportion of catch over length by Lady Hammond-WIIA from the candidate models (red solid line for the selected best model and blue dashed lines for other converged models), compared to the sample proportion of catch by length (gray dots for each paired tow within each station and black circles for the average across stations).

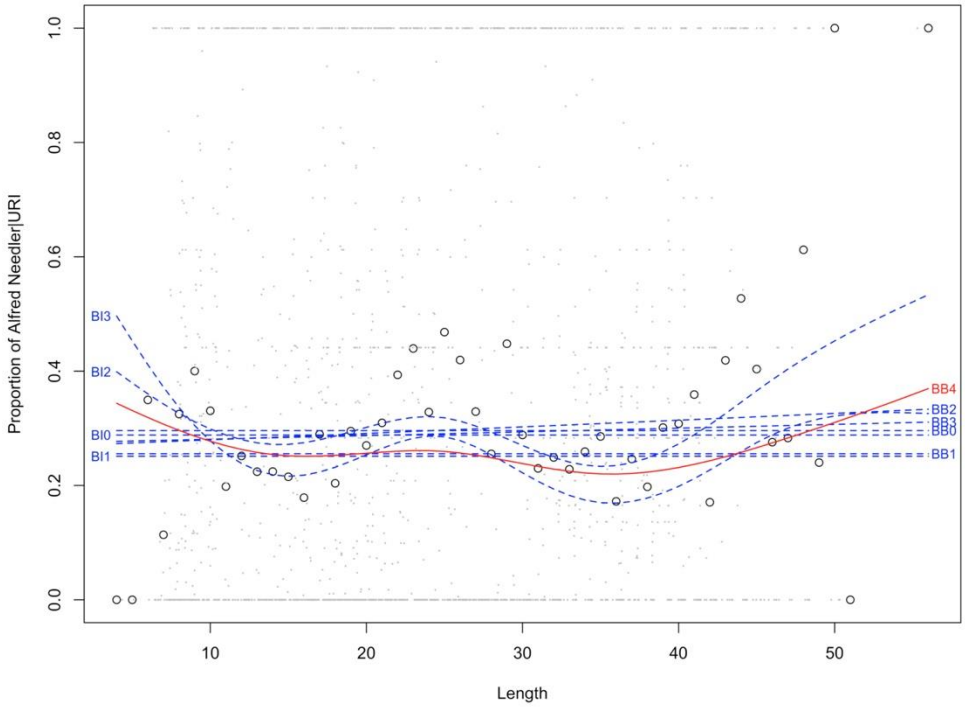

Figure 18: Comparative fishing analysis of NGSL 2004-2005, between *Alfred Needler*-URI and *Teleost*-Campelen: Estimated proportion of catch over length by *Alfred Needler*-URI from the candidate binomial and beta-binomial models (red solid line for the selected best model and blue dashed lines for other converged models), 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).

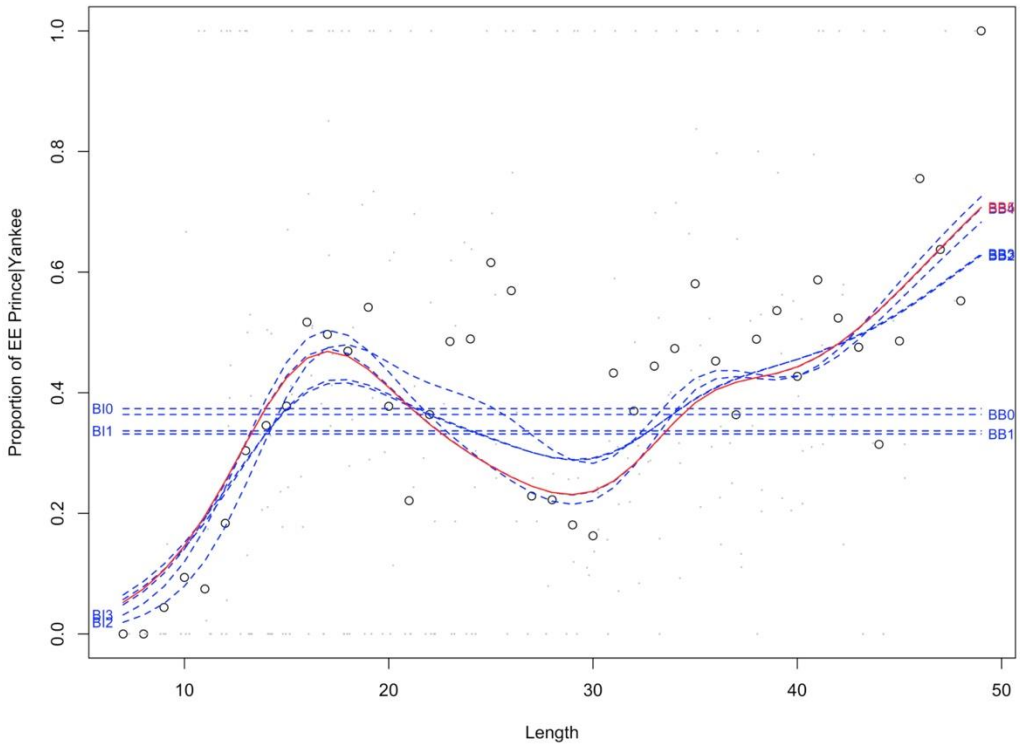

Figure 19: Comparative fishing analysis of SGSL 1985, between EE Prince-Yankee and Lady Hammond-WIIA: Estimated proportion of catch over length by EE Prince-Yankee from the candidate binomial and beta-binomial models (red solid line for the selected best model and blue dashed lines for other converged 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).

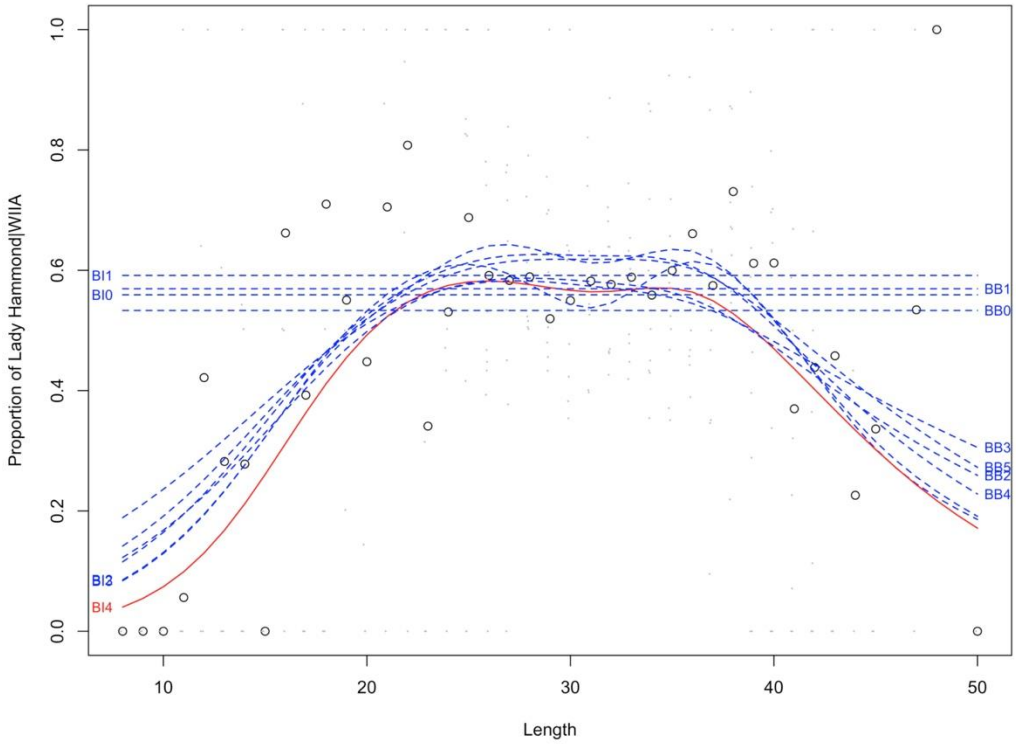

Figure 20: Comparative fishing analysis of SGSL 1992, between Lady Hammond-WIIA and *Alfred Needler*-WIIA: Estimated proportion of catch over length by Lady Hammond-WIIA from the candidate models (red solid line for the selected best model and blue dashed lines for other converged models), compared to the sample proportion of catch by length (gray dots for each paired tow within each station and black circles for the average across stations).

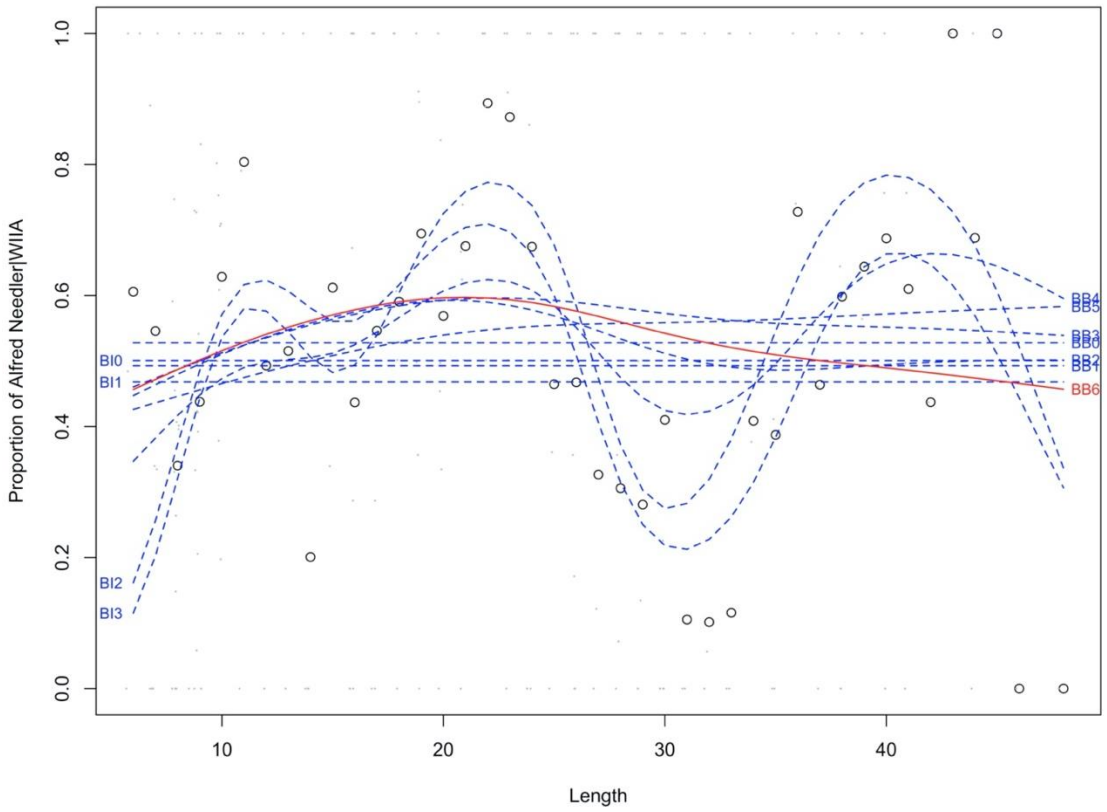

Figure 21: Comparative fishing analysis of SGSL 2004-2005, between *Alfred Needler*-WIIA and *Teleost*-WIIA: Estimated proportion of catch over length by *Alfred Needler*-WIIA from the candidate models (red solid line for the selected best model and blue dashed lines for other converged models), compared 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).

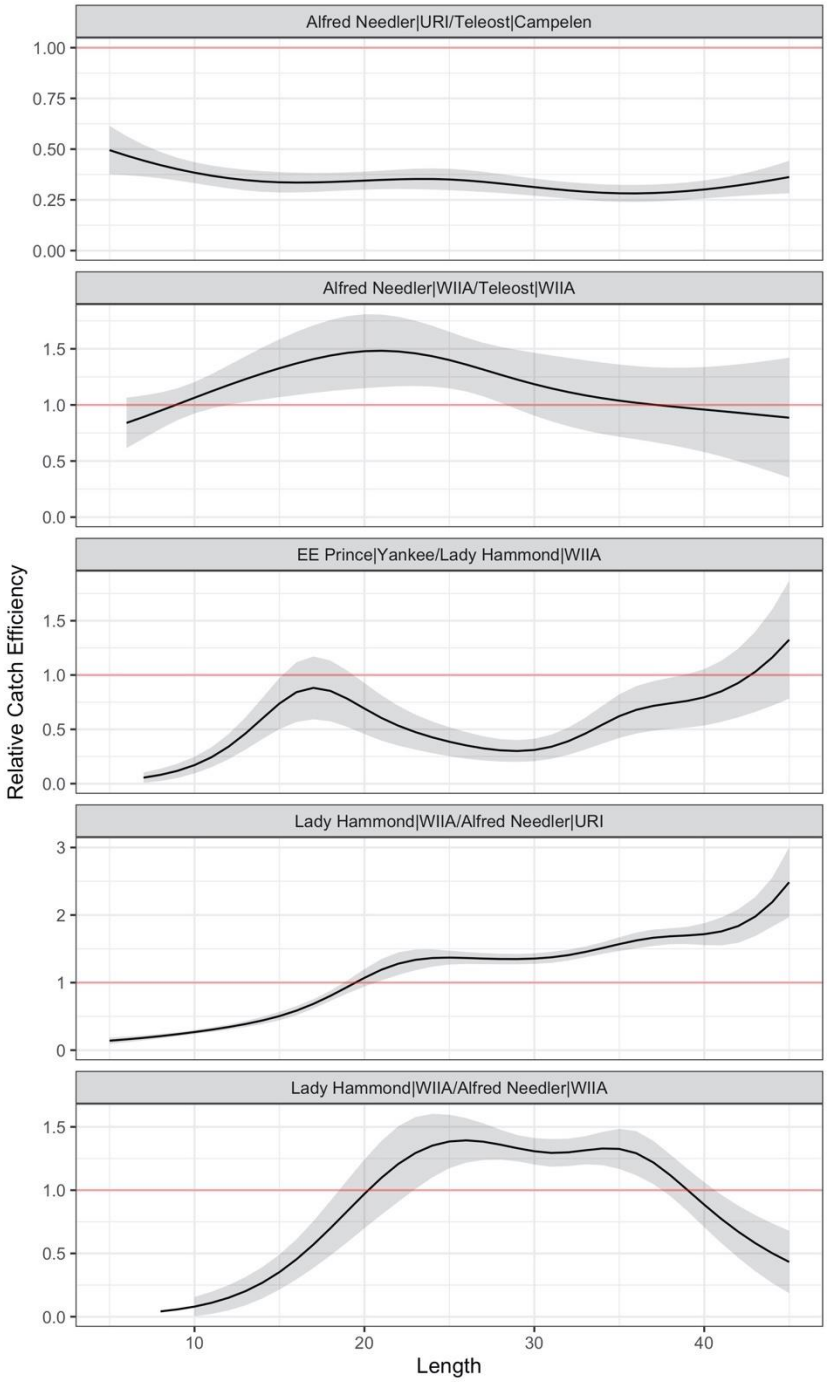

Figure 22: Estimated relative catch efficiency as a function of length from each comparative fishing analysis (black line), with one standard deviation (grey band). The red line represents a relative catch efficiency of 1, indicating no difference between the pair of vessel-gears.

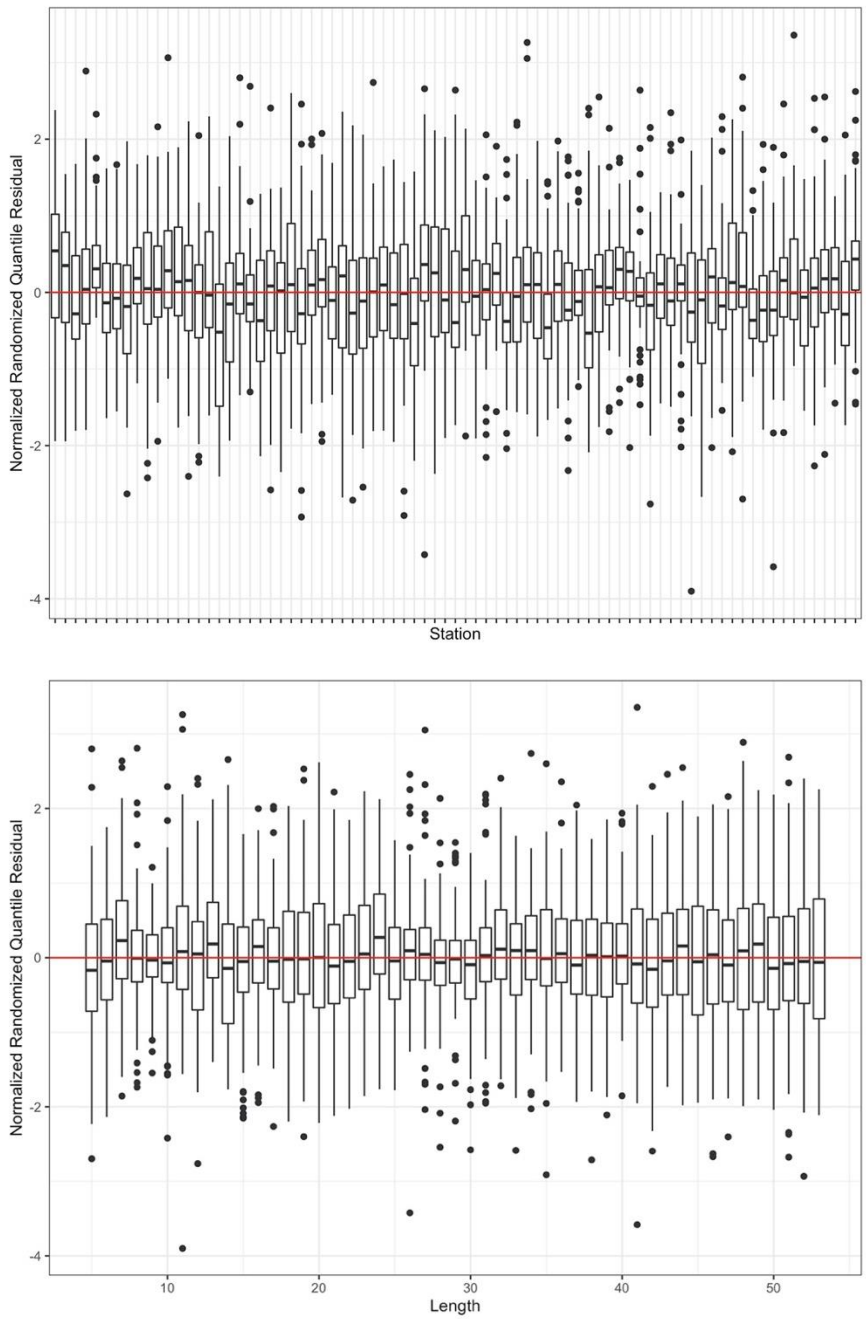

Figure 23: Comparative fishing analysis of NGSL 1990, between Lady Hammond-WIIA and *Alfred Needler*-URI: normalized randomized quantile residuals for each station (top panel) and for each length bin (bottom panel). The boxes indicate the 25% and 75% quantiles and the segment is the median; whiskers extends from the hinge to the largest/smallest values no further than 1.5 * IQR from the hinge (where IQR is the distance between the first and third quartiles).

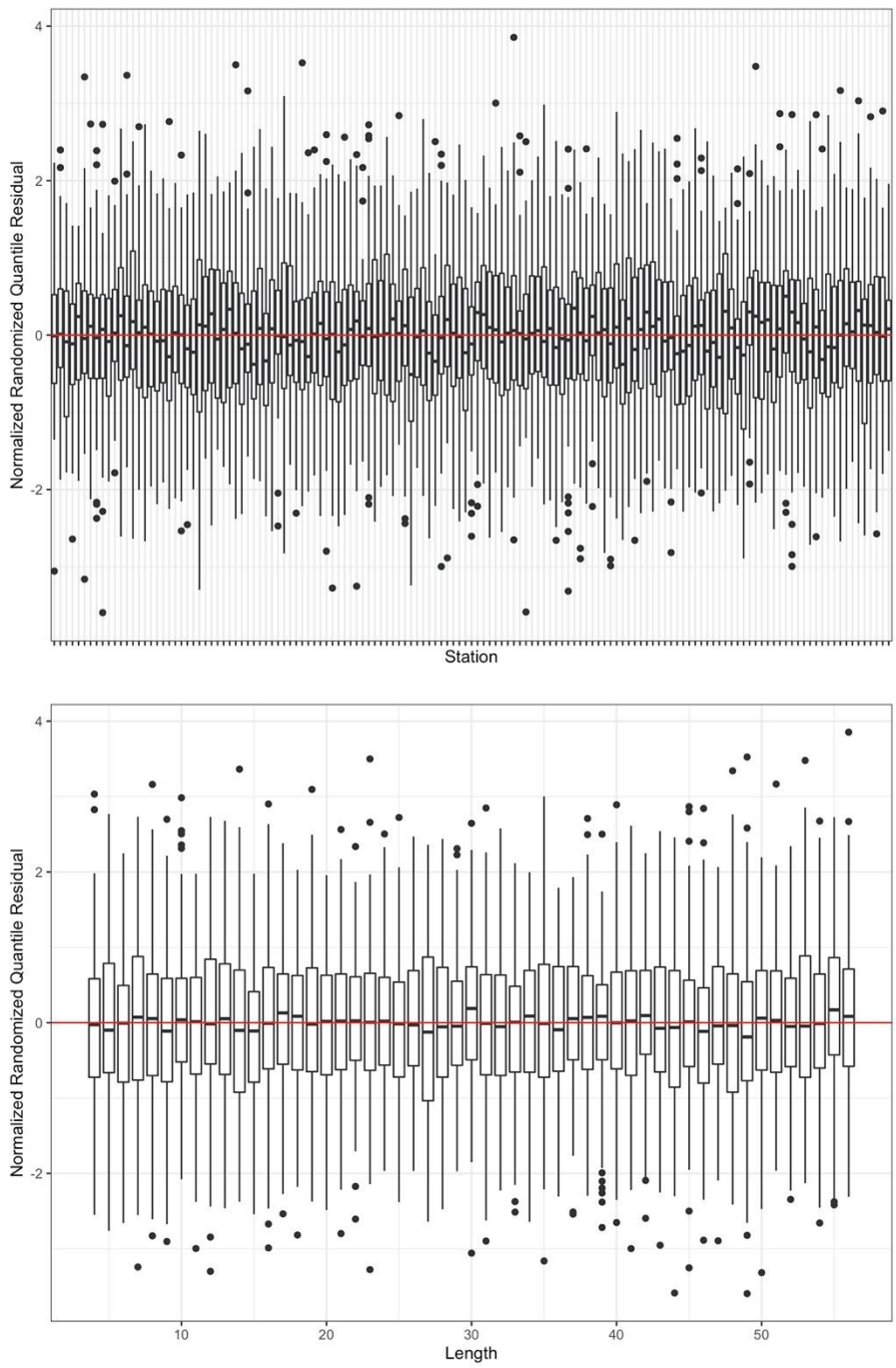

Figure 24: Comparative fishing analysis of NGSL 2004-2005, between *Alfred Needler*-URI and *Teleost*-Campelen: normalized randomized quantile residuals for each station (top panel) and for each length bin (bottom panel). The boxes indicate the 25% and 75% quantiles and the segment is the median; whiskers extends from the hinge to the largest/smallest values no further than 1.5 * IQR from the hinge (where IQR is the distance between the first and third quartiles).

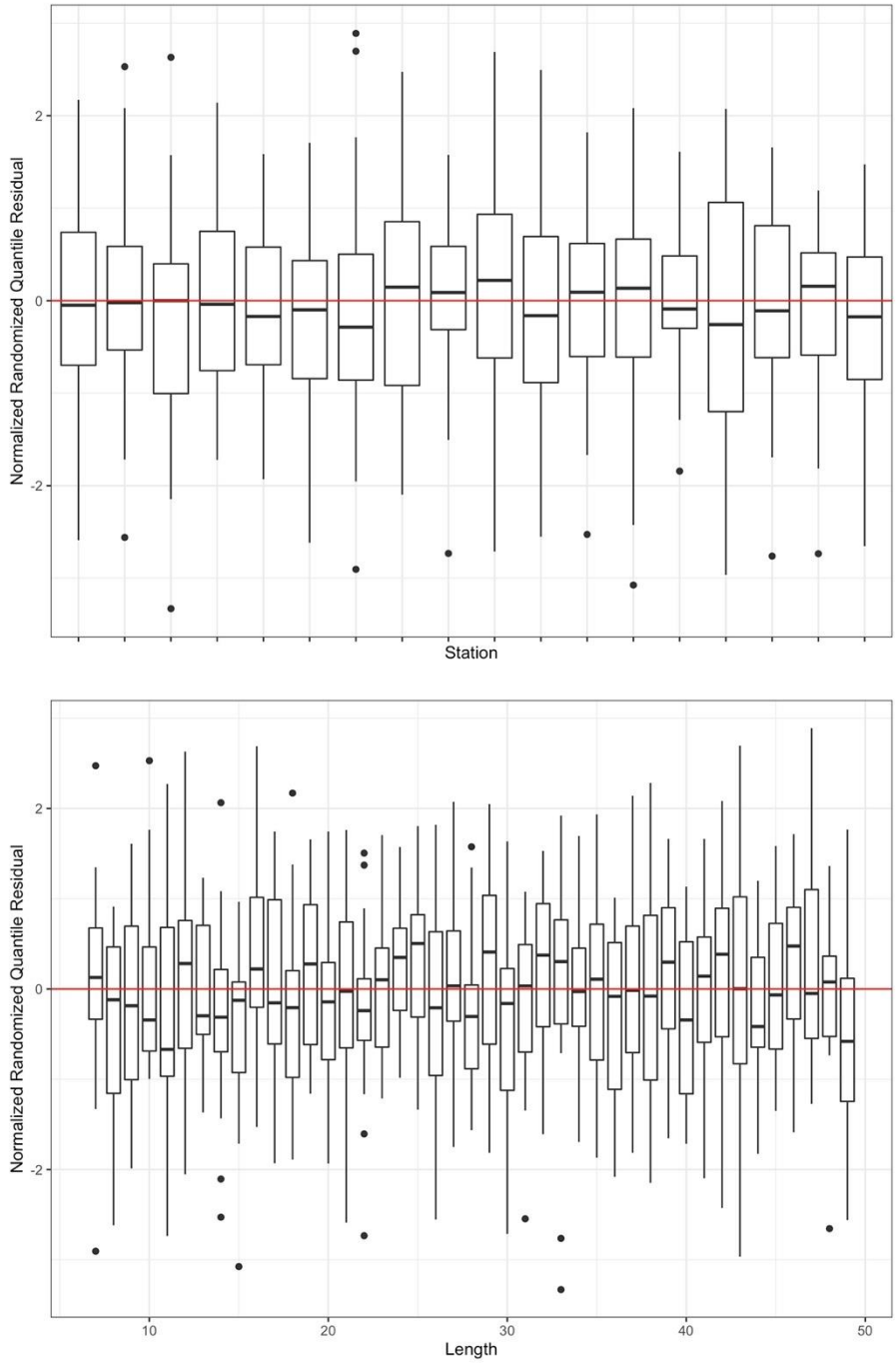

Figure 25: Comparative fishing analysis of SGSL 1985, between EE Prince-Yankee and Lady Hammond-WIIA: normalized randomized quantile residuals for each station (top panel) and for each length bin (bottom panel). The boxes indicate the 25% and 75% quantiles and the segment is the median; whiskers extends from the hinge to the largest/smallest values no further than 1.5 * IQR from the hinge (where IQR is the distance between the first and third quartiles).

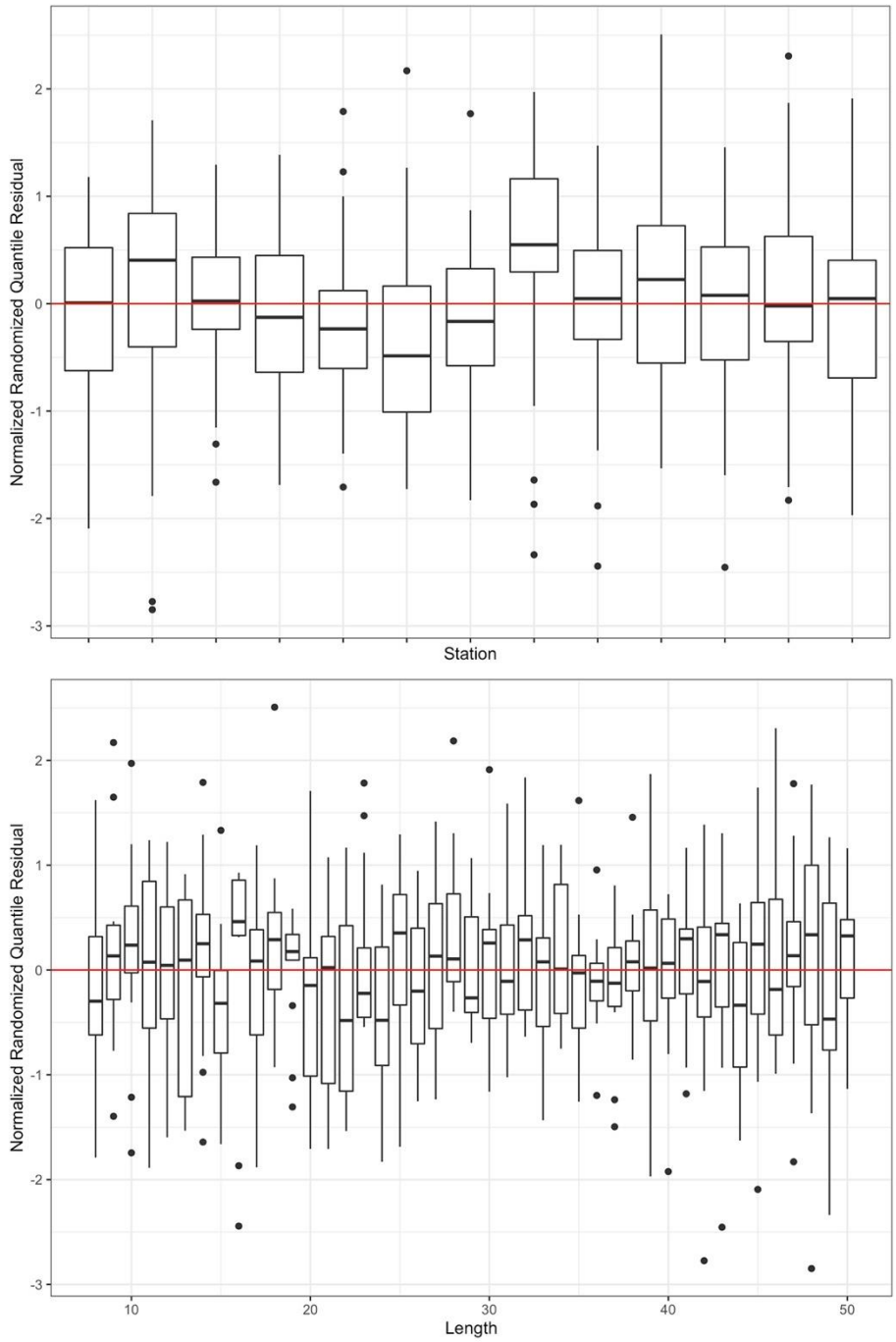

Figure 26: Comparative fishing analysis of SGSL 1992, between Lady Hammond-WIIA and *Alfred Needler*-WIIA: normalized randomized quantile residuals for each station (top panel) and for each length bin (bottom panel). The boxes indicate the 25% and 75% quantiles and the segment is the median; whiskers extends from the hinge to the largest/smallest values no further than 1.5 * IQR from the hinge (where IQR is the distance between the first and third quartiles).

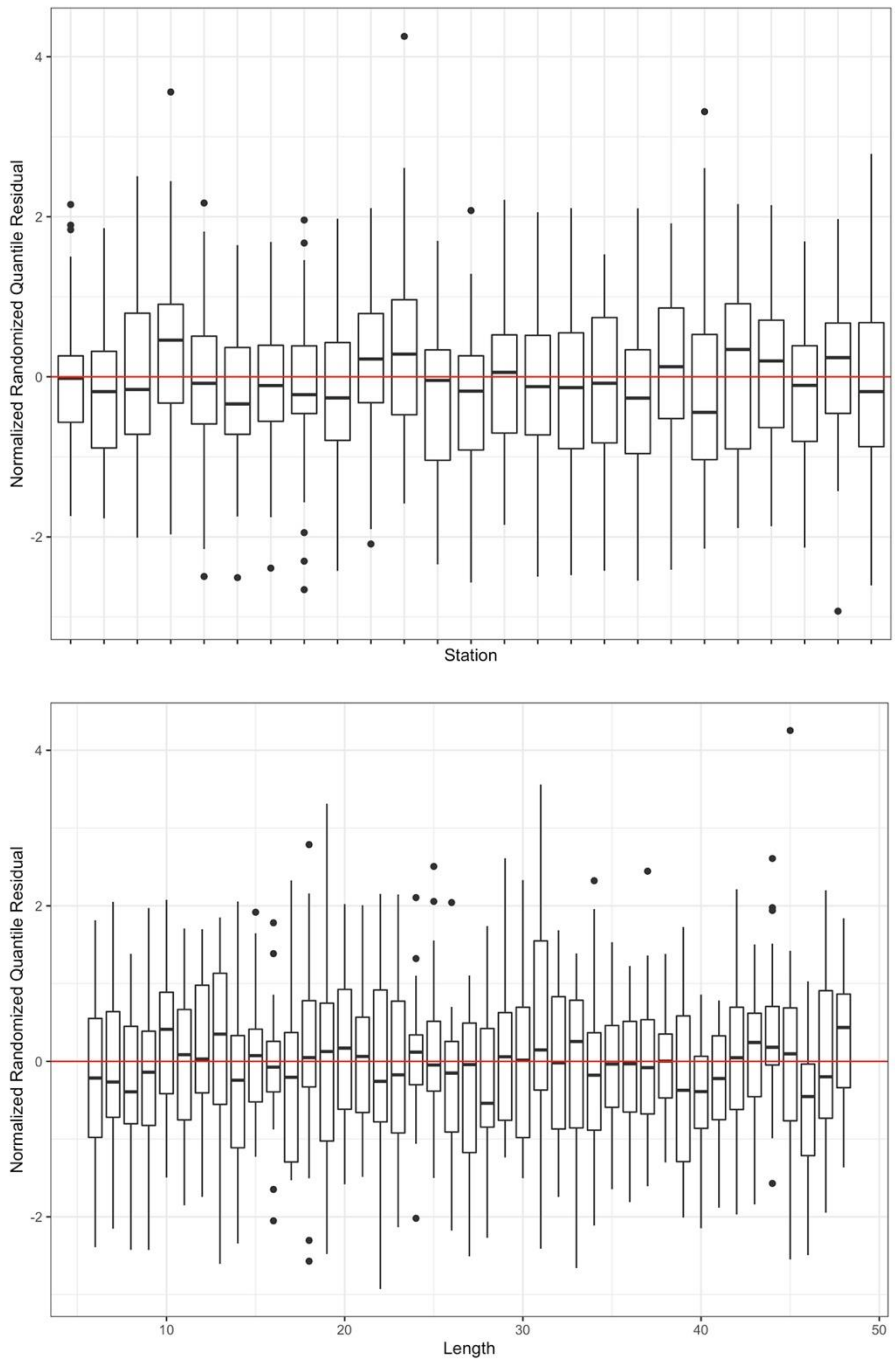

Figure 27: Comparative fishing analysis of SGSL 2004-2005, between *Alfred Needler*-WIIA and *Teleost*-WIIA: normalized randomized quantile residuals for each station (top panel) and for each length bin (bottom panel). The boxes indicate the 25% and 75% quantiles and the segment is the median; whiskers extends from the hinge to the largest/smallest values no further than 1.5 * IQR from the hinge (where IQR is the distance between the first and third quartiles).

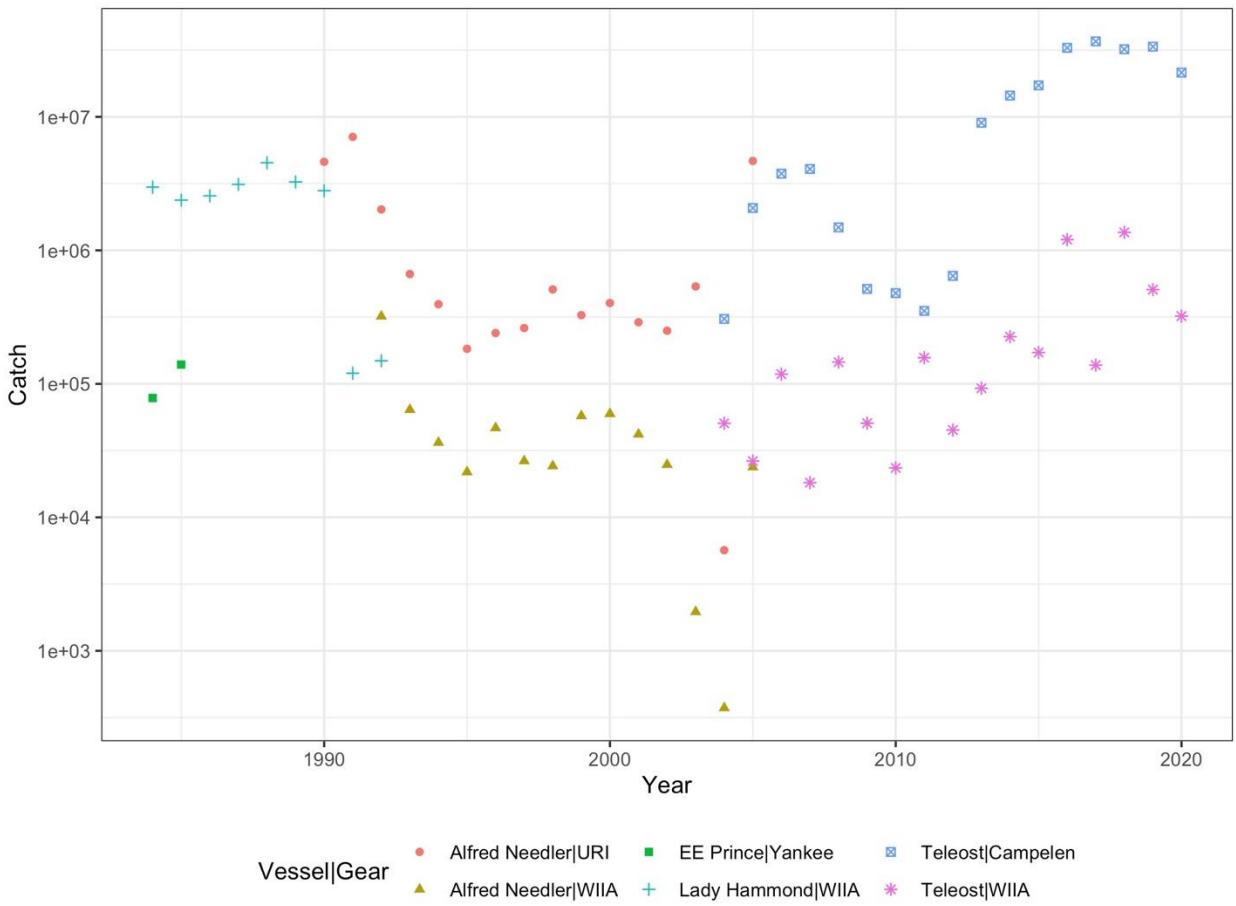

Figure 28: Annual total catch of redfish (effort-standardized catch numbers) by each vessel and gear using RV survey data from both northern and southern Gulf of St. Lawrence in the area of overlap. Catch levels have surged since 2013 as indicated by *Teleost*-Campelen.

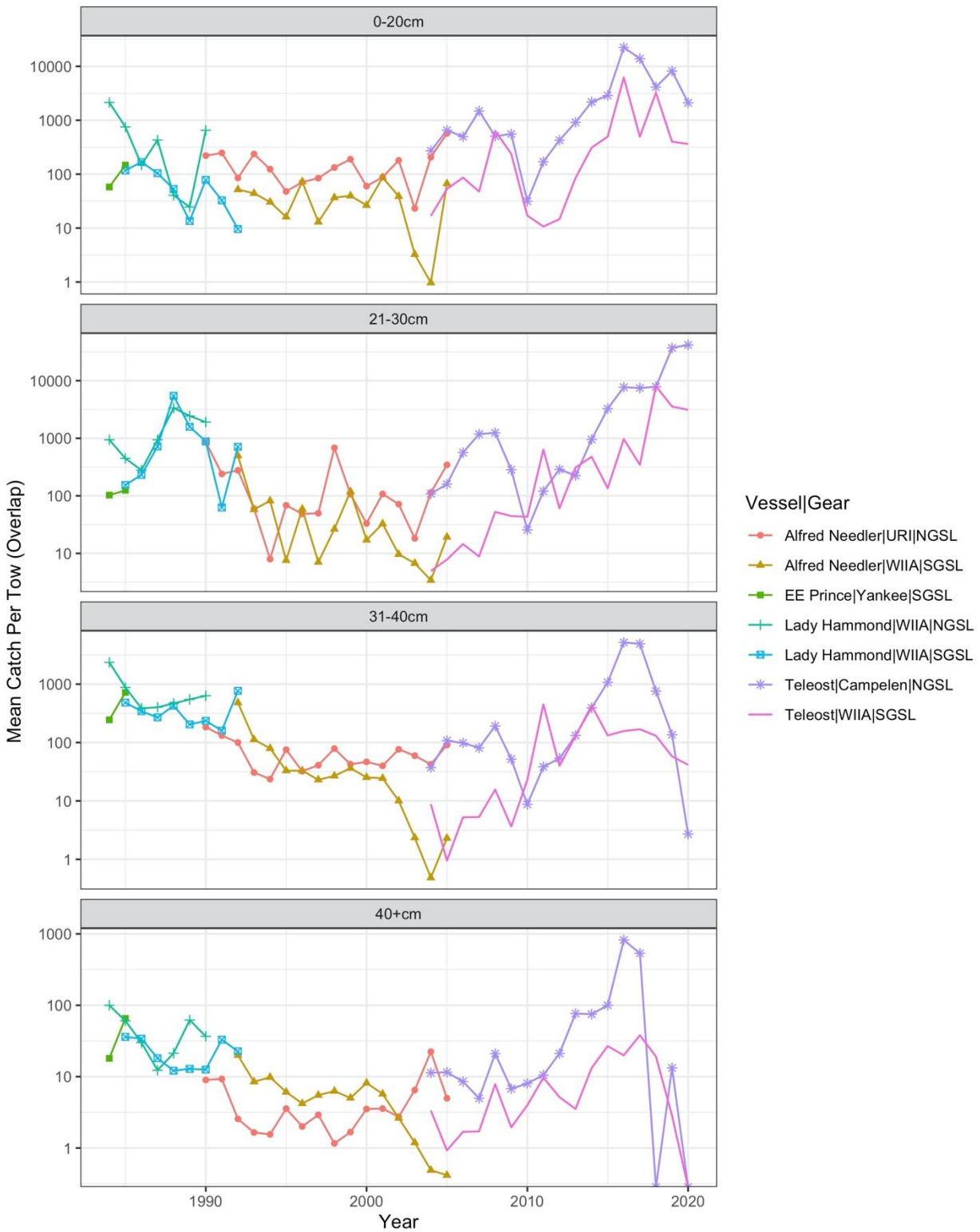

Figure 29: Time series of mean catch per tow for each vessel-gear and for four length groups using survey catches in the overlap area between the northern and southern Gulf of St. Lawrence (NGSL and SGSL).

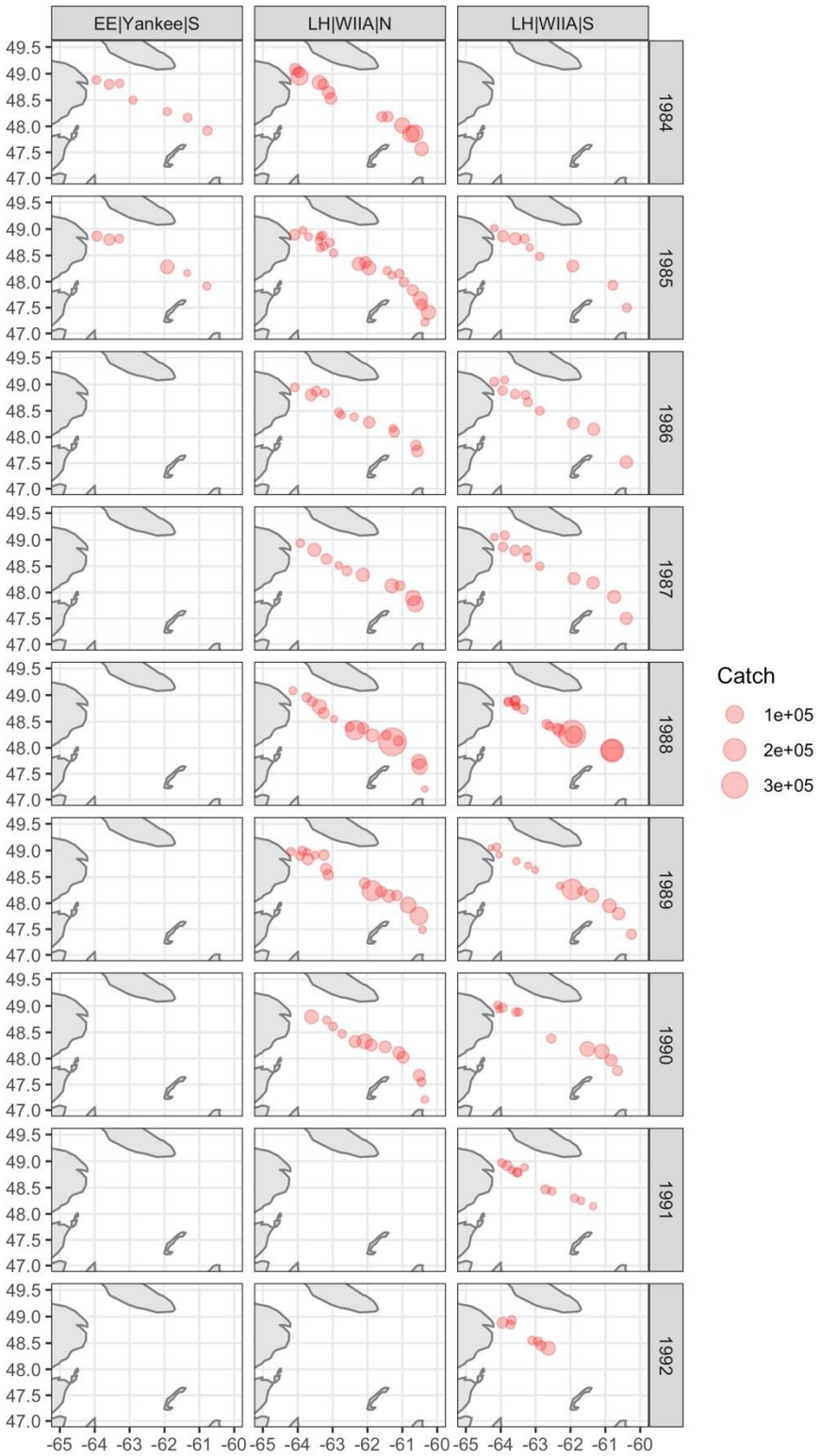

Figure 30: Spatial distribution of redfish survey catches (effort-adjusted) in the overlap area by each vessel-gear during 1984-1992 for a comparison between E.E. Prince(EE)-Yankee and Lady Hammond(LH)-WIIA in both the northern Gulf of St. Lawrence (denoted by N in the figure) and southern Gulf of St. Lawrence (denoted by S).

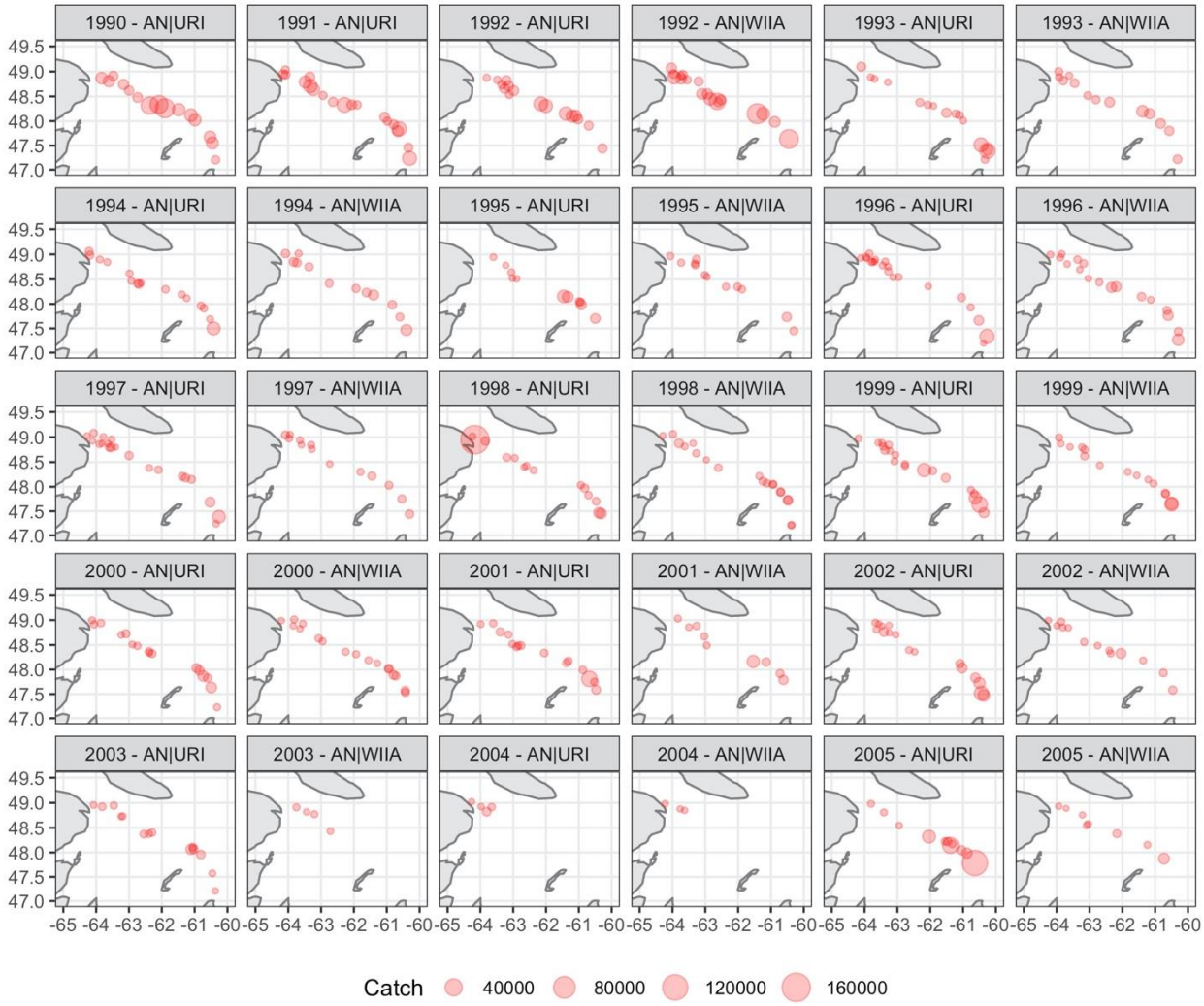

Figure 31: Spatial distribution of redfish survey catches (effort-adjusted) in the overlap area by each vessel-gear during 1990-2005 for a comparison between *Alfred Needler*(AN)-WIIA in the southern Gulf of St. Lawrence and *Alfred Needler*-URI in the northern Gulf of St. Lawrence. Note that 2003-AN|WIIA was in fact undertaken by the *CCGS W. Templemen*, sister ship to the *CCGS A. Needler*.

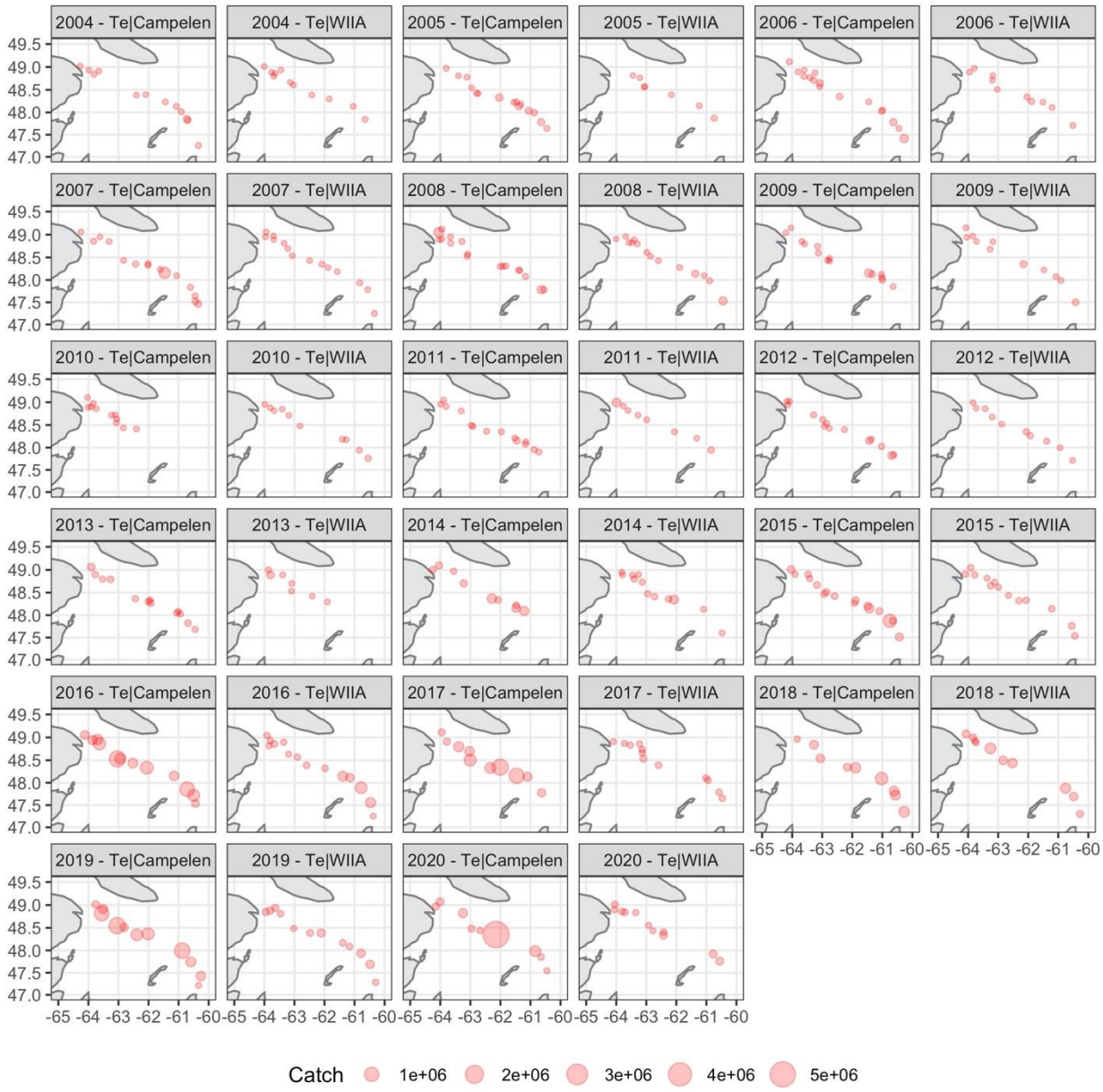

Figure 32: Spatial distribution of redfish survey catches (effort-adjusted) in the overlap area by each vessel-gear during 2004-2020 for a comparison between *Teleos*t(Te)-WIIA in the southern Gulf of St. Lawrence and Teleost-Campelen in the northern Gulf of St. Lawrence.

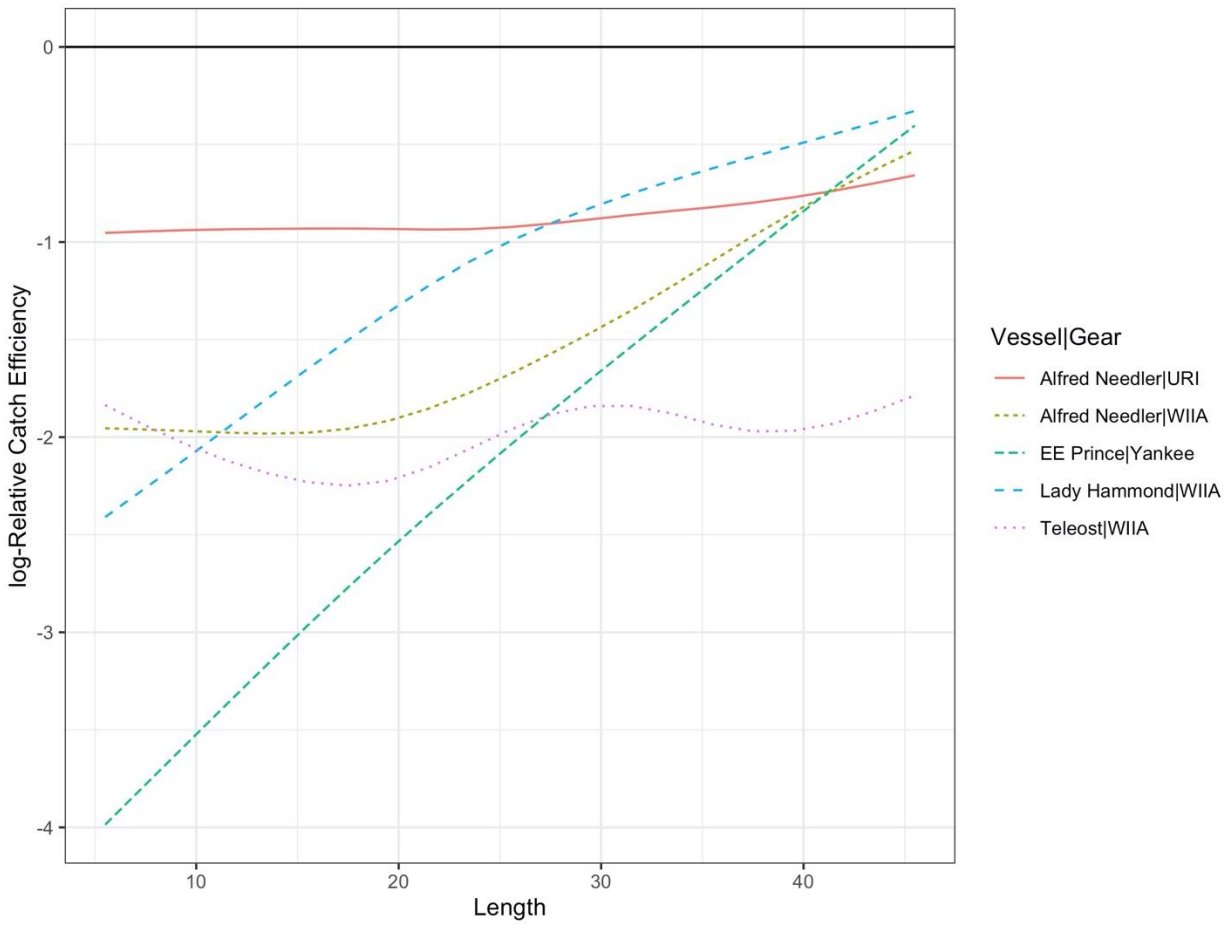

Figure 33: Estimated log-relative catch efficiency for each gear (all gears calibrated to Teleost-Campelen equivalent, in black horizontal line).

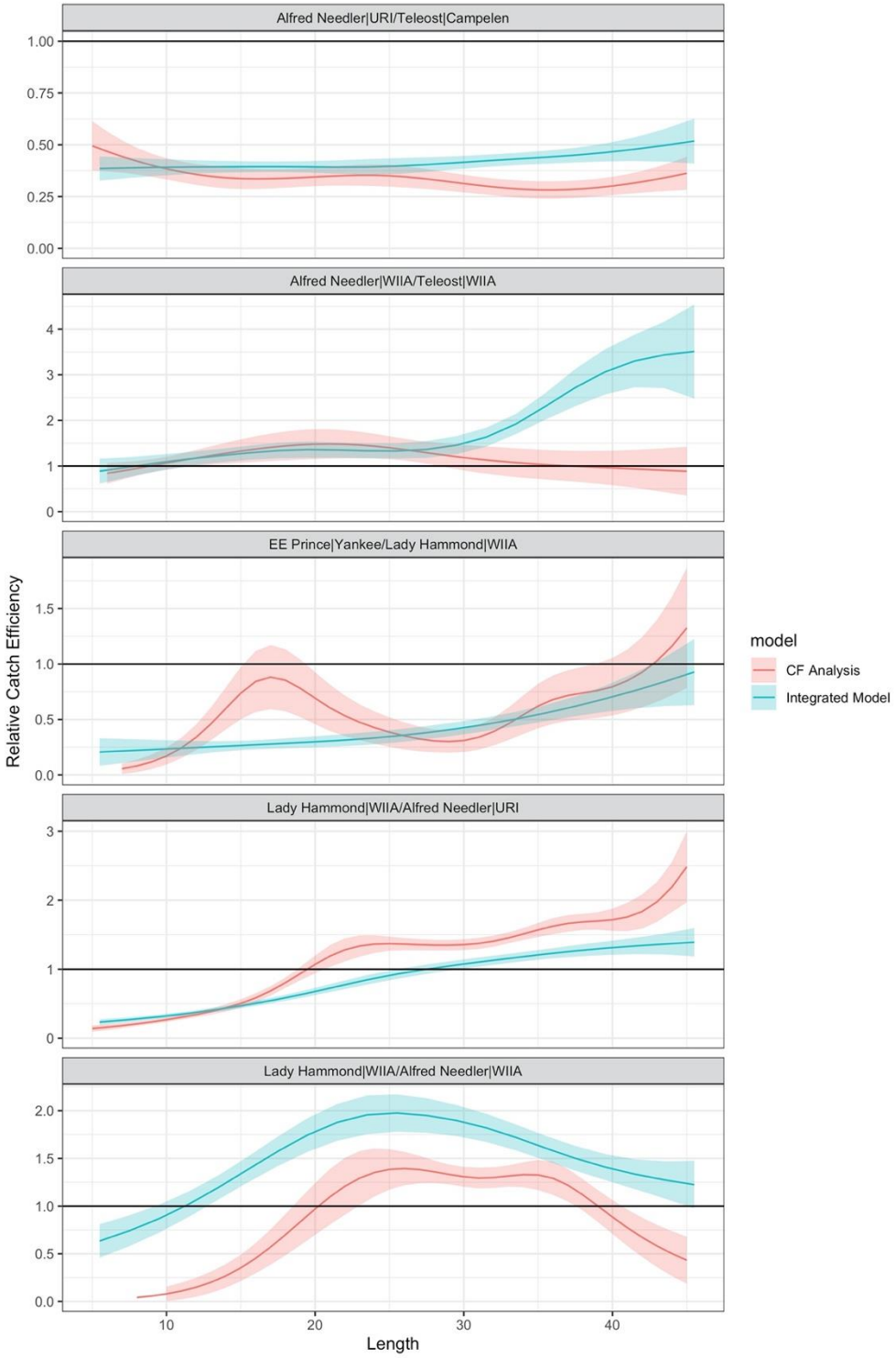

Figure 34: Estimated catch efficiencies with one standard deviation estimated from the integrated model and transformed to be pairwise corresponding to the five comparative fishing experiments, in comparison with results from the separate comparative fishing (CF) analyses. Estimates are plotted in red and blue lines for the two models, respectively, and the bands are standard deviations.

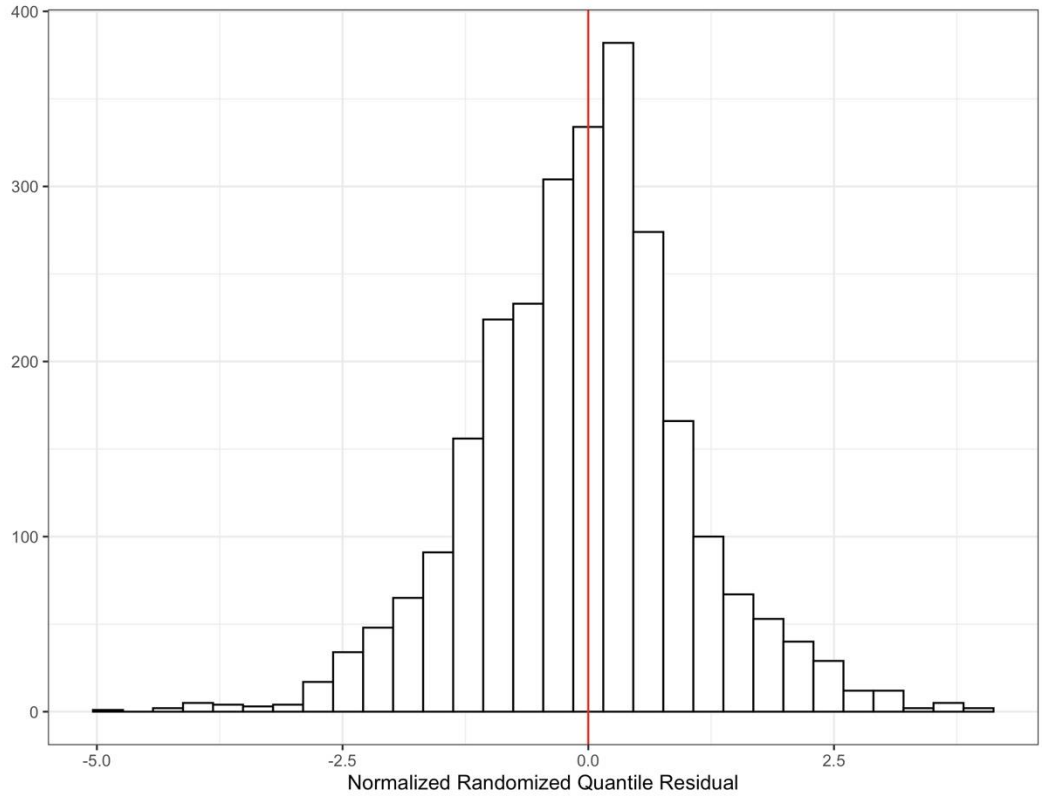

Figure 35: Residual diagnostics for paired catches: normalized randomized quantile residuals were calculated for each pair and each length bin based on the beta-binomial distribution and are generally compared to a gaussian distribution.

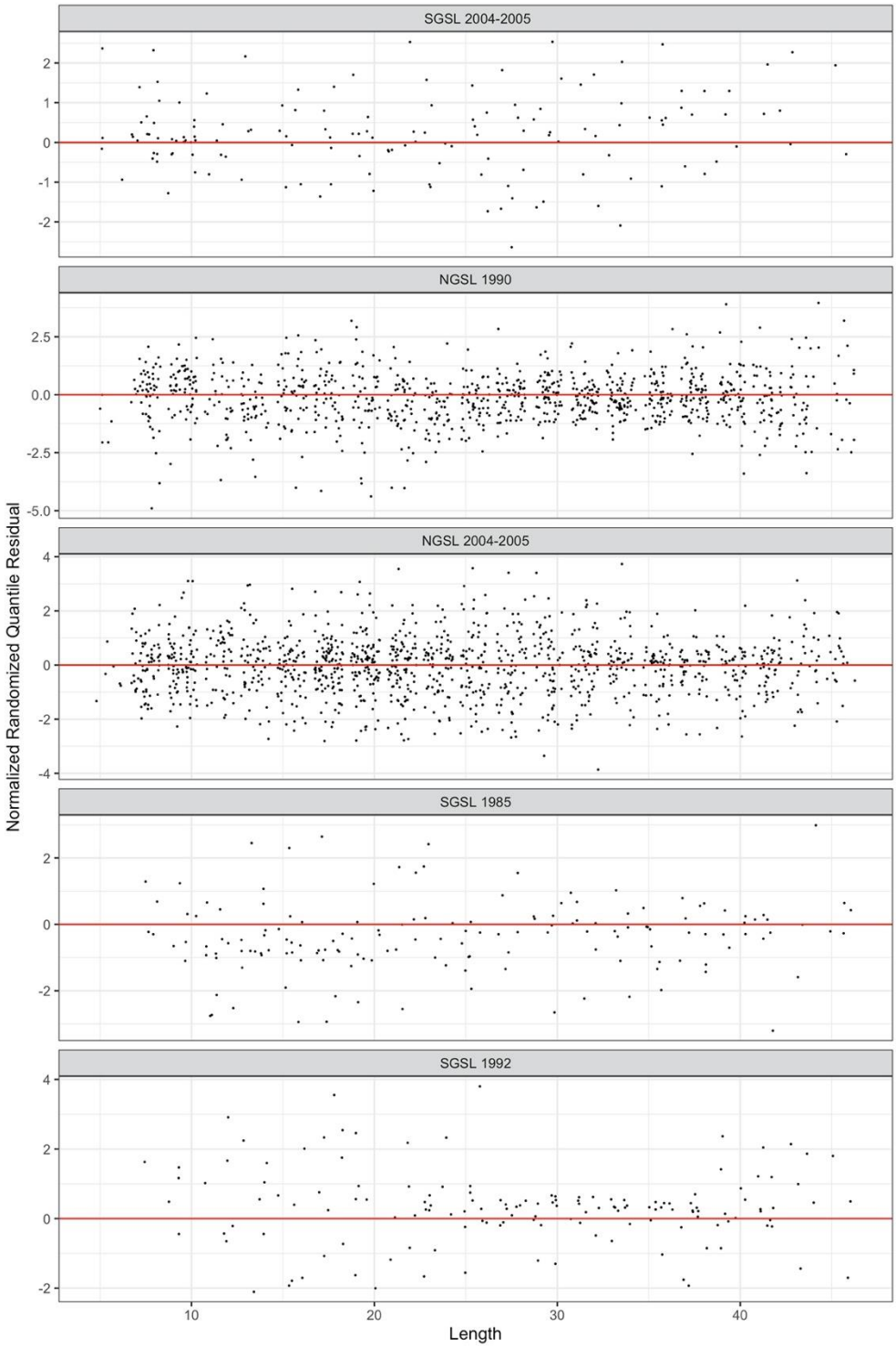

Figure 36: Residual diagnostics for paired catches: randomized quantile residuals were checked for each comparative fishing experiment and each length bin (black dots, jittered to enhance clarity).

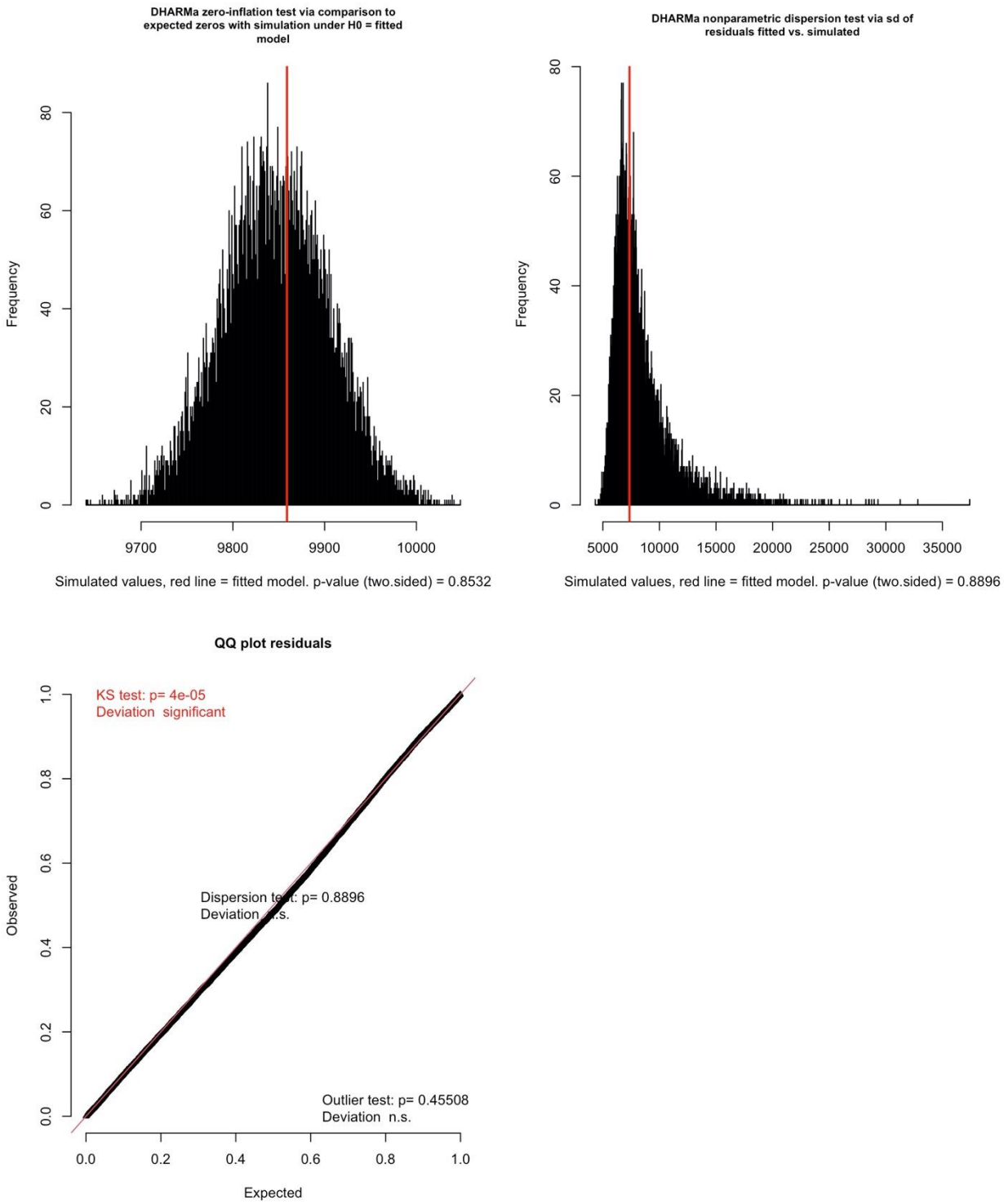

Figure 37: *DHARMa* residual diagnostics for the catch distribution assumption in the integrated analysis, including tests for zero-inflation, over-dispersion and goodness-of-fit of the negative binomial distribution. The model past all tests except the KS-test, indicating some deviation from the overall distributional assumption.

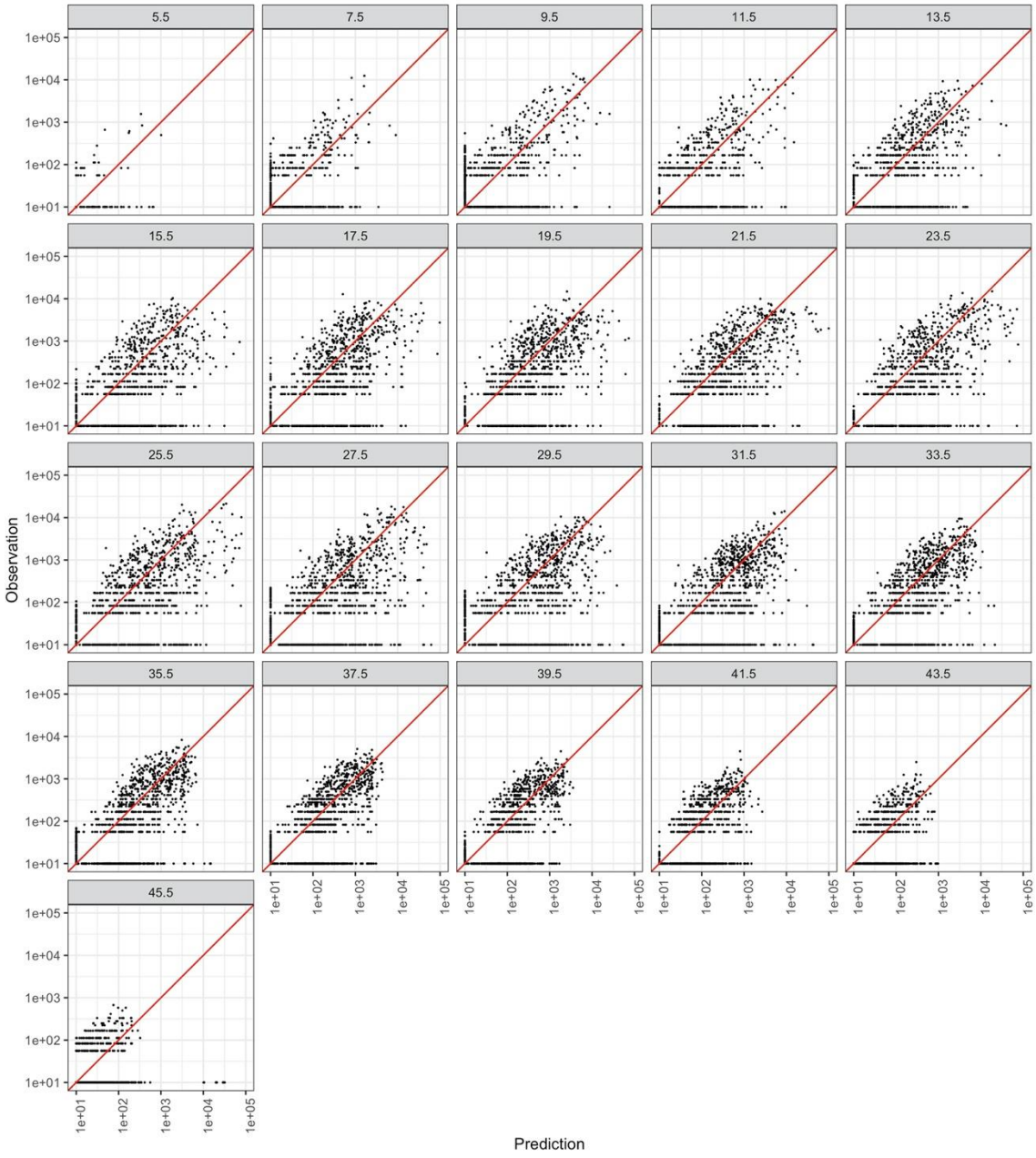

Figure 38: Predicted versus observed catches for each 2cm length group (any quantities under 10 were conformed to 10 exactly in order to preserve all data as well as the log scale of the axes in the plot). The name of each panel indicates the median of the length group, e.g., 5.5 indicates the length group 5-6cm.

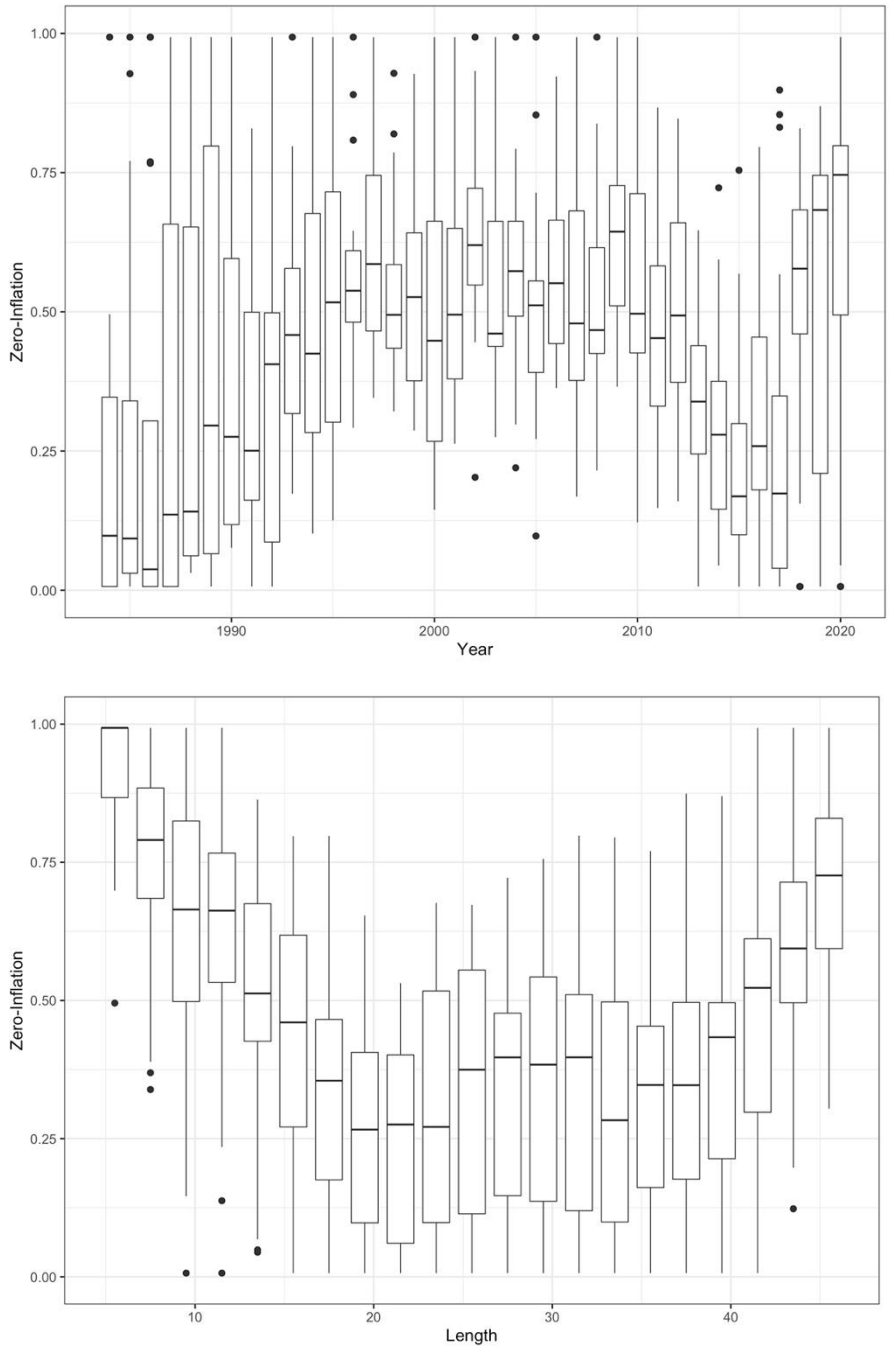

Figure 39: Annual trend (upper panel) and length relationship (lower panel) of the estimated zero-inflation rate, or proportion of true zeros in the overlap area.

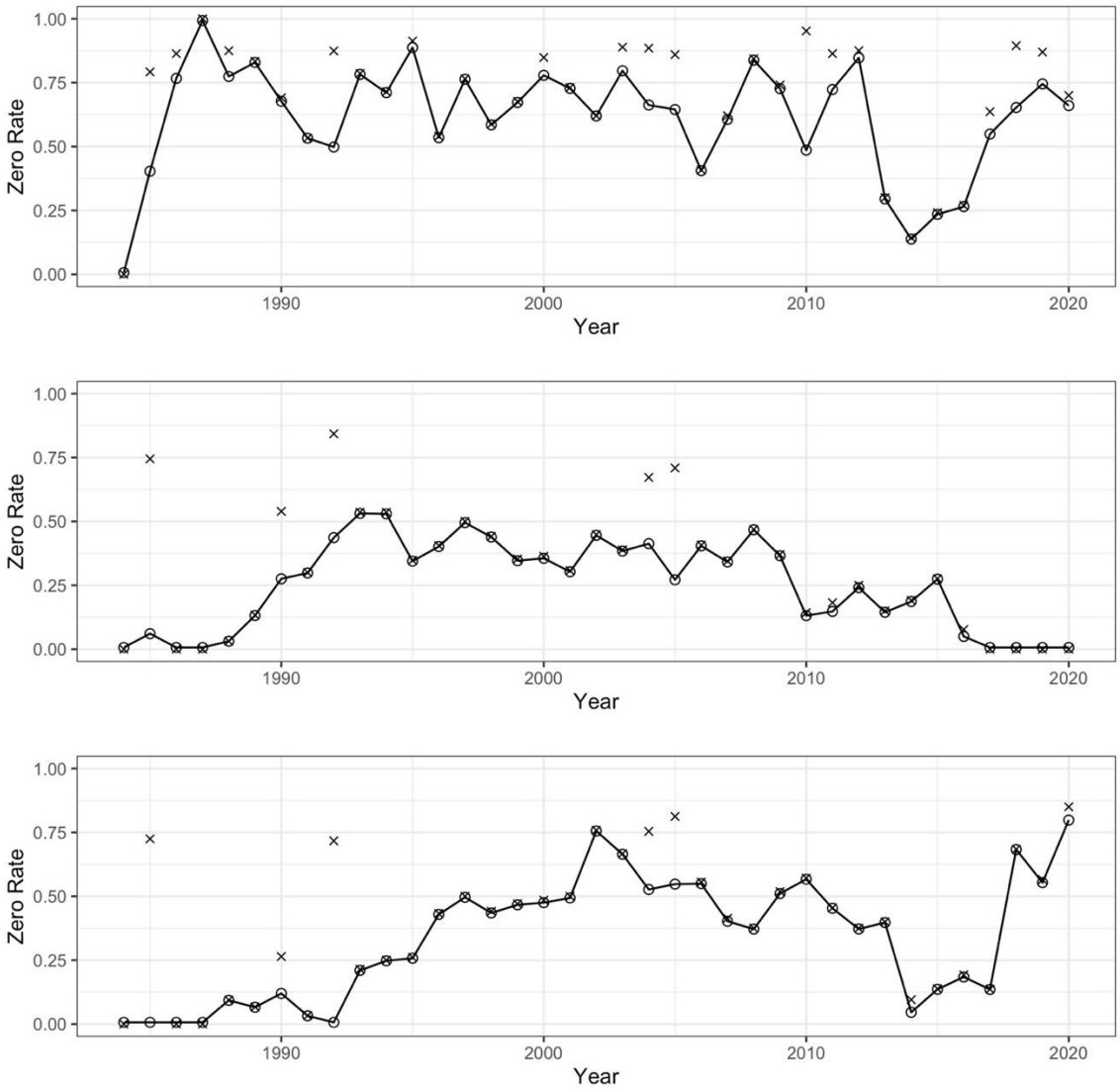

Figure 40: Sample proportion of zero catches (crosses) for the three length groups chosen for illustration: 11-12cm, 21-22cm, and 31-32cm (from top to bottom), and their corresponding estimated zero-inflation rate (in circles), or proportion of true zeros. The zero-inflation rate (zip) is an indicator of redfish presence probability (1-zip) within the area.

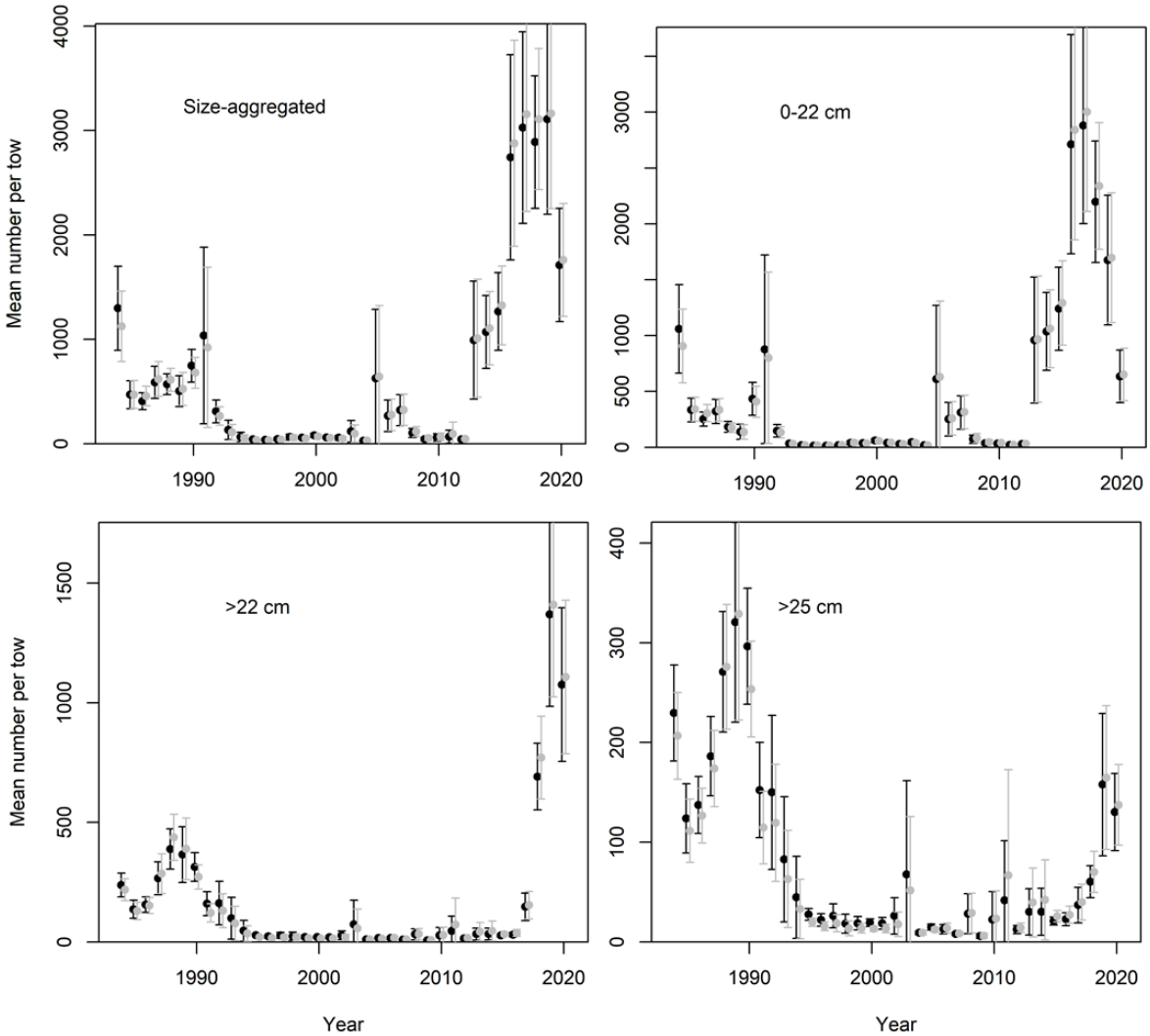

Figure 41: Standardized redfish abundance indices (mean numbers per tow, with 95% confidence interval) for the entire GSL with standardization based on comparative fishing only (black points) and from the integrated analysis (grey points) for all sizes and by size groups used in the assessment of Units 1+2 redfish.

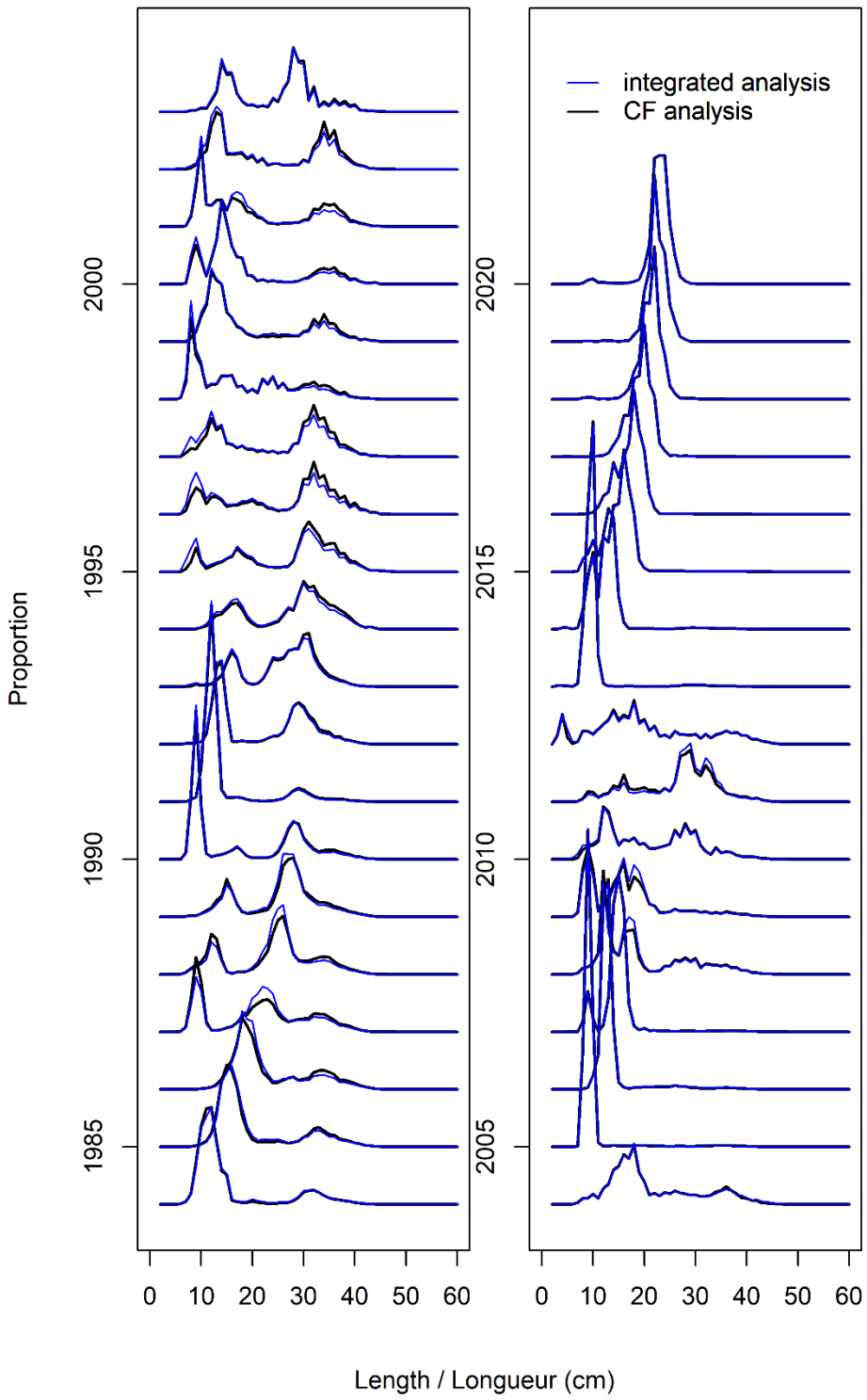

Figure 42: Annual relative length frequencies (proportions at length) for the standardized redfish abundance indices for the entire GSL, with standardization based on comparative fishing only (black lines) and from the integrated analysis (blue lines).