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

Canadian Science Advisory Secretariat (CSAS)

Research Document 2023/038

Arctic Region and Ontario and Prairie Region

Development of Spatial Operating Models to Test Survey Design and Calibrate a New Survey Index for Northwest Atlantic Fisheries Organization Subarea 0+1 (offshore) Greenland Halibut (*Reinhardtius hippoglossoides*)

Quang C. Huynh and Tom Carruthers

Blue Matter Science
2150 Bridgman Avenue
North Vancouver, British Columbia V7P 2T9

May 2023

Canada

Foreword

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

Published by:

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

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

ISSN 1919-5044
ISBN 978-0-660-48808-0 Cat. No. Fs70-5/2023-038E-PDF

Correct citation for this publication:

Huynh, Q.C., and Carruthers, T. 2023. Development of Spatial Operating Models to Test Survey Design and Calibrate a New Survey Index for Northwest Atlantic Fisheries Organization Subarea 0+1 (offshore) Greenland Halibut (*Reinhardtius hippoglossoides*). DFO Can. Sci. Advis. Sec. Res. Doc. 2023/038. iv + 35 p.

Aussi disponible en français :

*Huynh, Q.C., et Carruthers, T. 2023. Élaboration de modèles opérationnels spatiaux pour tester la conception de relevé pour le flétan du Groenland (*Reinhardtius hippoglossoides*) du sous-zone 0+1 (au large des côtes) de l'Organisation des pêches de l'Atlantique Nord-Ouest. Secr. can. des avis sci. du MPO. Doc. de rech. 2023/038. iv + 38 p.*

TABLE OF CONTENTS

ABSTRACT

In 2018, the *Paamiut* research vessel used to conduct the bottom trawl survey in Baffin Bay and Davis Strait was retired, with no possibility to conduct comparative tows with the new *Tarajoq* research vessel. With the transition between vessels and trawl types generating a new index series starting in 2022, alternative methods that could be used to improve management advice for Greenland Halibut in the Northwest Atlantic Fisheries Organization Subarea 0+1 (offshore) is sought. This paper describes a framework for developing a spatial operating model used to simulate surveys and indices of abundance. Fishery and survey data were used to condition a population model and a spatiotemporal model to develop an operating model to describe historical abundance and distribution trends of Greenland Halibut (*Reinhardtius hippoglossoides*). Using the SimSurvey software package, the current random, depth-stratified survey design (in Divisions 0A-South and 1CD) was simulated and compared to an alternative, hypothetical survey that also included Division 0B. Simulation studies can identify the relative benefit of various components of survey design, e.g, additional survey coverage, strata boundaries, and set density, towards index precision within the logistical constraints of the survey, for example, ship time.

The exploratory nature of this work was presented to identify models for potential assessment of this stock in the future. These approaches can address the survey gap between 2018-2021 and subsequent vessel change in two ways. First, the spatiotemporal model could be used to calibrate the catchability between the RV *Paamiut* and RV *Tarajoq* which would allow stitching the indices of abundance from the two vessels into a single time series. Initial exploration of model-based calibration with the 2019 survey on the FV *Helga Maria* detected lower catchability in some depth strata and size classes of fish, consistent with field observations. Further simulation testing and validation of the approach is recommended prior to broader use. Second, retrospective simulation through SimSurvey of an index of abundance, using catchability and selectivity estimates of the RV *Paamiut* and RV *Tarajoq*, can demonstrate if the two surveys can hypothetically provide similar inference on stock trends and magnitude over the same time period.

1. INTRODUCTION

Indices of abundance from biological surveys are valuable data used to inform stock assessments and management of fish stocks. Surveys use consistent sampling protocols to avoid preferential sampling of the stock and obtain unbiased estimates of stock trends over time. These indices can be used to fit a population assessment model and/or inform a harvest control rule for catch advice.

Management of the offshore component of Northwest Atlantic Fisheries Organization (NAFO) Subarea 0+1 Greenland Halibut (*Reinhardtius hippoglossoides*) has been informed by an index of abundance time series developed from the trawl survey in Divisions 0A-South (south of 72º N) and 1CD. The *Paamiut* research vessel used for the trawl survey was slated for retirement and Fisheries and Oceans Canada (DFO) was planning to request additional ship time to conduct comparative tows with the new vessel and gear in order to provide empirical data for calibrating catchability. Unfortunately, early retirement of the RV *Paamiut* was necessitated in 2018 with no ability to perform the comparative trawl sets. The *Helga Maria* fishing vessel was used for the 2019 survey, although catchability differences were observed with the gear that precluded the use of the 2019 survey to update the index time series (Nogueira and Treble 2020). The new research vessel, the *Tarajoq*, conducted the 2022 survey and will be used in the future. However, the intervening gap between the RV *Paamiut* and RV *Tarajoq* surveys (2018-2022) creates an information gap until a new index time series with the RV *Tarajog* is established.

For the short and medium term (up to 5 years), alternative methods are needed to inform management due to the survey gap. Recent development of software facilitating spatial modeling of fish populations provide an opportunity to develop spatial operating models and perform simulations to evaluate survey design. While there is currently no accepted stock assessment for NAFO Subarea 0+1 (offshore) Greenland Halibut, there is sufficient fishery-dependent and fishery-independent data to develop operating models that explore hypotheses regarding stock abundance, with age-structured models, and the spatial distribution, with spatiotemporal models. Alternative sampling designs or catchability assumptions that reflect use of different vessel and gear types can then be simulated and evaluated through performance measures against the operating model.

These models can then be used to address the survey gap in two ways. First, spatiotemporal models can accommodate survey data from several vessels in an attempt to estimate calibration factors in the absence of comparative towing. Second, retrospective simulation of index time series using catchability and selectivity estimates of the RV *Paamiut* and RV *Tarajoq* can demonstrate if the two surveys hypothetically provide similar inference on stock trends and magnitude over the same time period.

This paper describes a framework for development of a spatial operating model and simulation for Greenland Halibut using various open source software packages, including SimSurvey (Regular et al. 2020), sdmTMB (Anderson et al. 2021), and openMSE (Hordyk et al. 2021). SimSurvey is the focal package used in the simulation framework, with the other supporting packages used to inform inputs required in SimSurvey. This paper demonstrates a comparison of two indices of abundance from surveys which differ in spatial coverage. A schematic of the simulation is provided in Figure 1. A demonstration of the model-based calibration with the sdmTMB spatiotemporal model using the 2019 survey data is provided in the appendices. Then a description of the approach with retrospective simulation is presented as future work in the discussion.

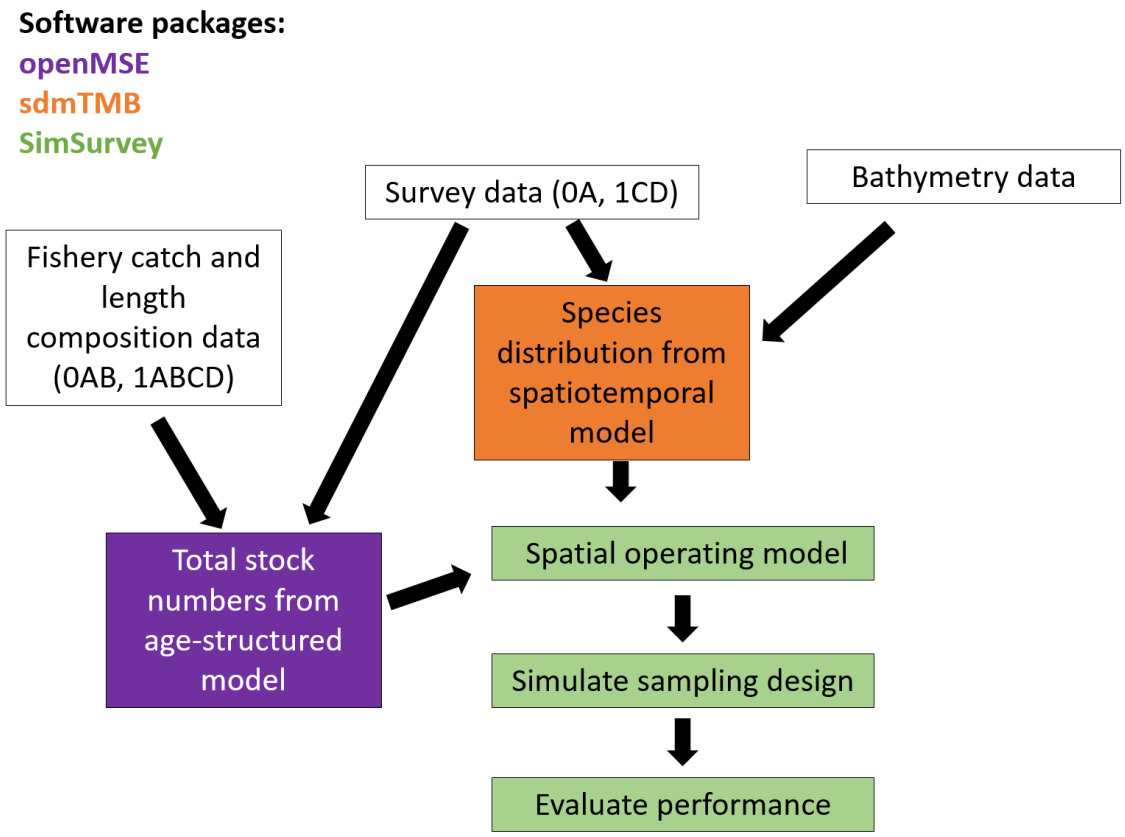

Figure 1. Schematic drawing of the simulation design that outlines the data inputs (white boxes) into the corresponding software packages (colored boxes).

2. METHODS

2.1. AGE-STRUCTURED POPULATION MODEL

The operating model for SimSurvey was developed in two steps. First, total stock abundance ($N_{t,a}$ by year *t* and age *a*) are needed to set up the population. Estimates of abundance were obtained using the Rapid Conditioning Model (RCM) implemented in the SAMtool package (Huynh et al. 2021), a component of the openMSE software ecosystem (Hordyk et al. 2021). RCM is a single-area, age-structured population model designed for conditioning operating models for management strategy evaluation (see Description at openMSE).

Catches (1987–2019) from Divisions 0AB and 1ABCD and the index of abundance series from the trawl survey in Divisions 0A (south of 72° N) and 1CD were used (Treble and Nogueira 2020). The index was developed from stratified total abundance estimates for years (1999, 2001, 2004, 2008, 2010, 2012, 2014-2017) in which both 0A and 1CD were sampled using the RV *Paamiut* (i.e., the 2019 survey was excluded). The RCM was also fitted to the corresponding length frequencies from both the fishery and survey (Figures 2–4). Three fishery fleets (trawl, gillnet, and longline) were modeled in RCM as the length composition indicated that they caught different size classes of fish.

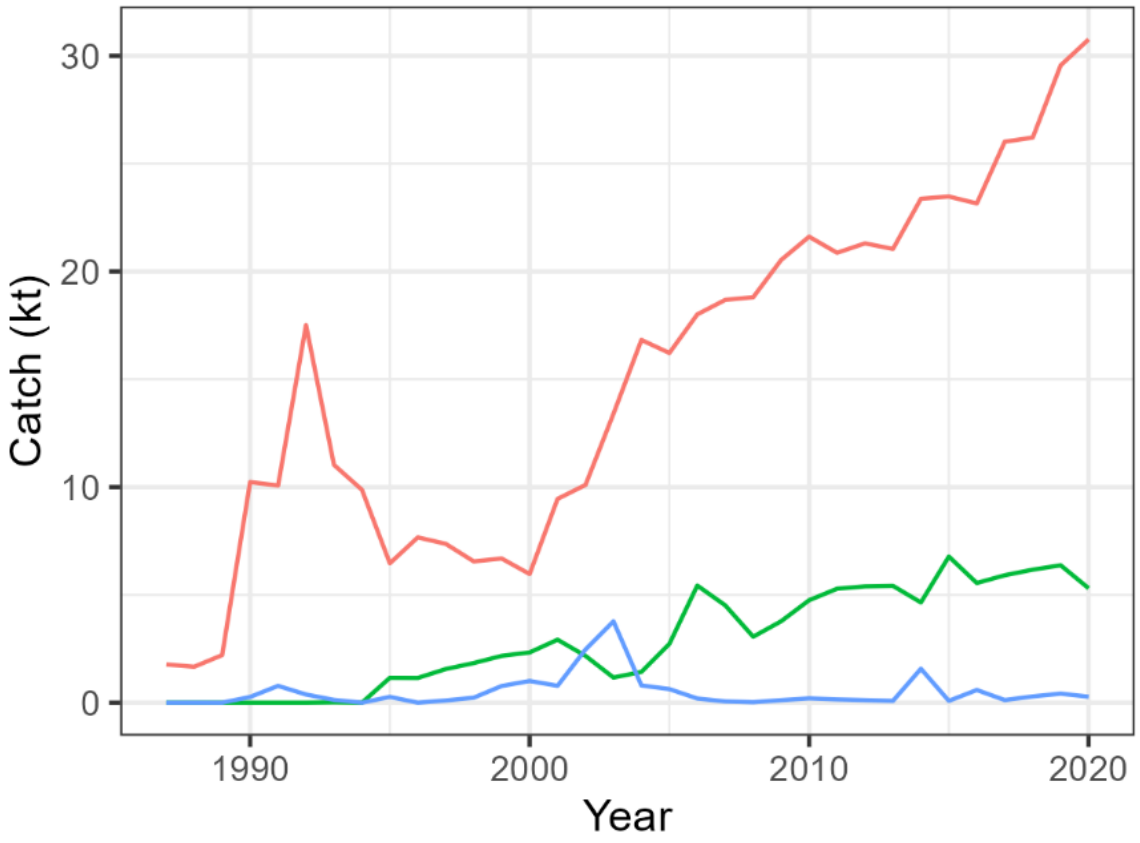

Figure 2. Greenland Halibut fishery catches (kt) from the trawl (red line), gillnet (green line), and longline (blue line) fisheries in Divisions 0AB and 1A-D from 1987 to 2020.

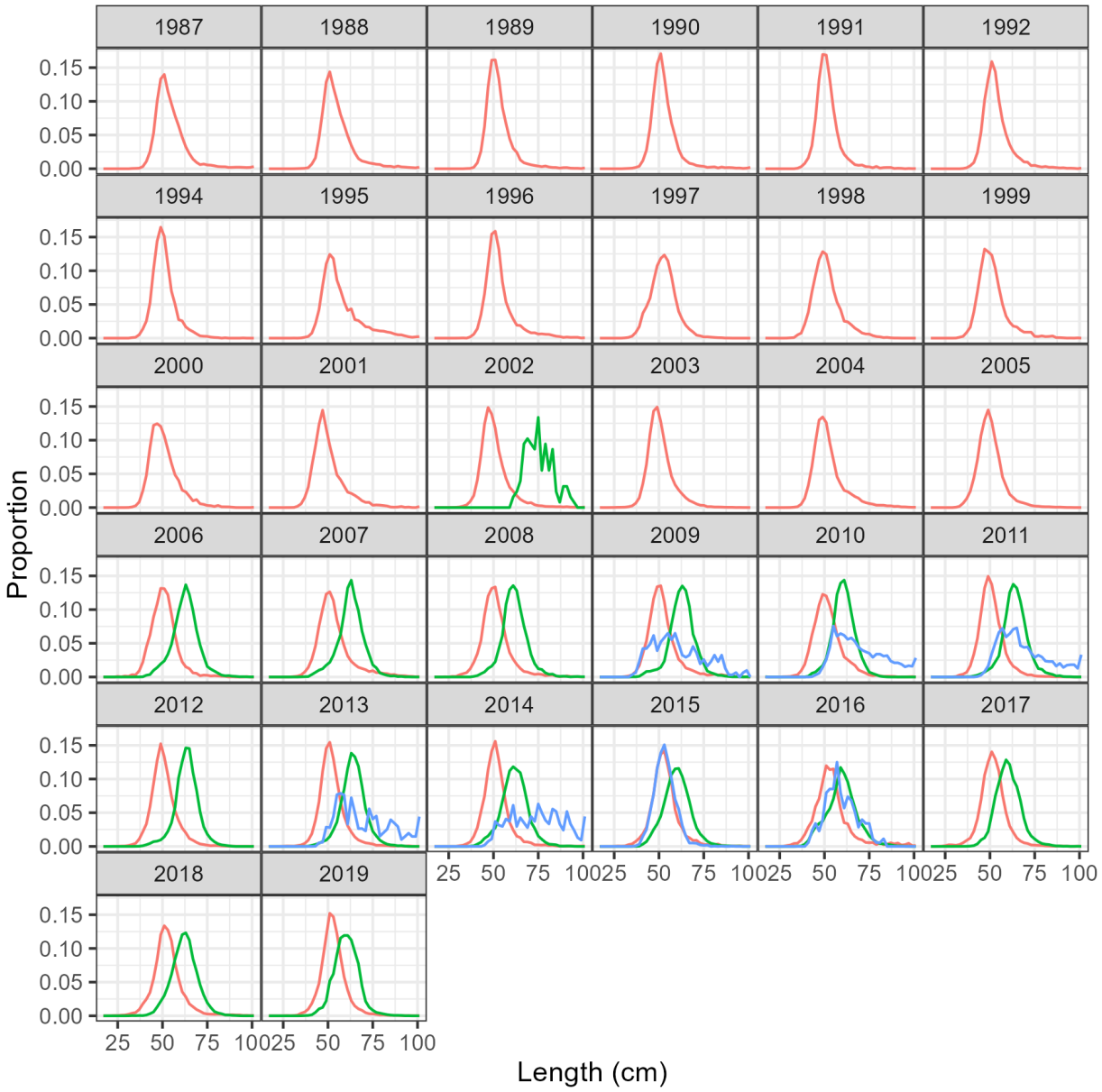

Figure 3. Greenland Halibut length composition from the trawl (red line), gillnet (green line), and longline (blue line) fisheries from 1987-2019 from Divisions 0AB and 1ABCD. Most longline length composition were collected from Norway. However, the 2015 longline composition does not include Norway catches and only include catch from Canada.

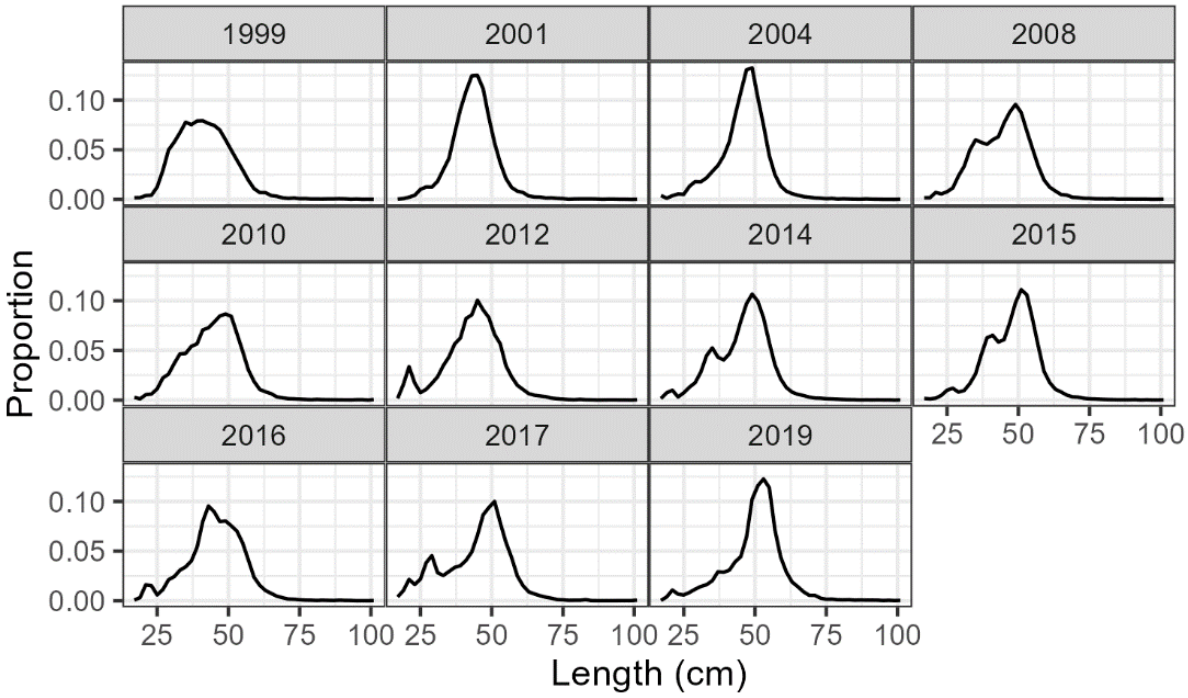

Figure 4. Greenland Halibut length composition from the trawl survey combined index (in Divisions 0A-South and 1CD) from 1999-2019.

Biological parameters, such as growth and maturity, were estimated from survey age and length samples (Table 1, Figures 5–7). Sex-specific parameters were estimated from length-at-age and maturity-at-length. Female parameters were used as it is assumed that population fecundity is constrained by female abundance and that most large fish observed are female (sex-specific data from the fishery were not available for this analysis). No sexual dimorphism was seen in weight-at-length so a single set of parameters were estimated from data from both sexes. Natural mortality was estimated to be 0.11 using the Then et al. (2015) estimator. This value was within the range of values used in the assessment for the southern Greenland Halibut stock (Subarea 2, Divisions 3KLMNO; Morgan et al. 2019).

Table 1. Biological parameters used in the population model. L_∞ is the von Bertalanffy asymptotic length, K is the von Bertalanffy growth coefficient, a_0 is the theoretical age of length zero, τ is the coefficient of variation in length at age, α and β are length-weight conversion parameters, L_{50} and L_{95} are the length of 50% and 95% maturity, respectively.

Process	Function	Parameter Value
Length at age; $L(a)$	$L(a) = L_\infty[1 - \exp -K(a - a_0)]$	L_∞= 93.6 cm K = 0.055 a_0 = -2.2
Standard deviation in length at age; $\sigma(L(a))$	$\sigma(L(a)) = L(a) \times \tau$	$\tau = 0.11$

Process	Function	Parameter Value
Weight at length; $W(L)$	$W(L) = \alpha L^{\beta}$	$\log(\alpha) = -12.5$ $\beta = 3.23$
Maturity at length; $m(L)$	$m(L) = \left[1 + \exp\left(-\log(19)\left[\frac{L - L_{50}}{L_{95} - L_{50}}\right]\right)\right]^{-1}$	L_{50} = 61 cm L_{95} = 84 cm
Natural mortality; M	$M = 4.118\, K^{0.73} L_{\infty}^{-0.33}$	M = 0.11

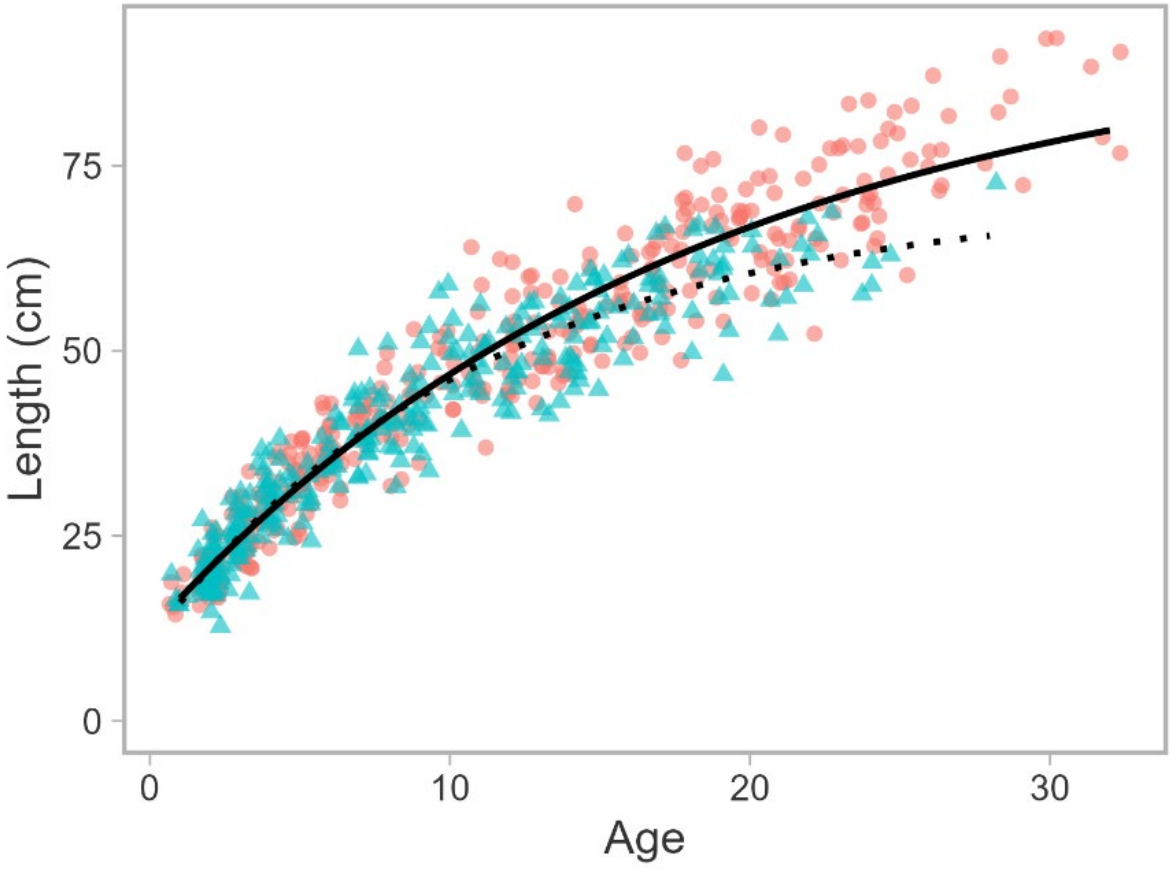

Figure 5. Length-at-age (females in red circles, males in blue triangles) of Greenland Halibut. Only the biological samples collected in the trawl survey in 2014 and 2017 have been aged to date. Lines indicate the predicted mean length-at-age (females in solid line, male in dotted line) from fitting the von Bertalanffy growth function.

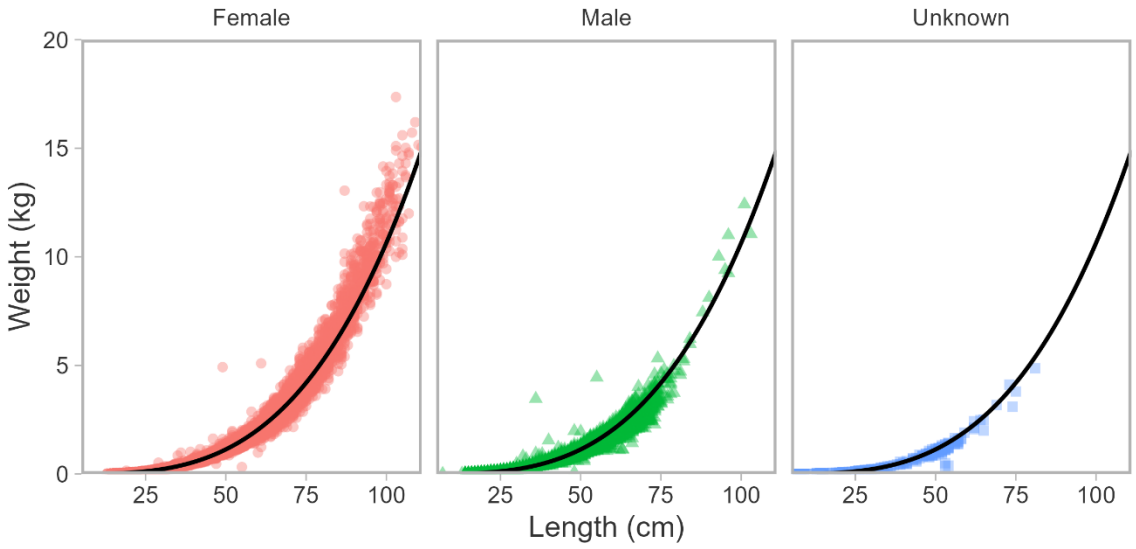

Figure 6. Weight-at-length (females in red circles, males in green triangles, unsexed in blue squares) of Greenland Halibut from biological samples collected from the trawl survey from 1999-2019. The black line indicates the predicted weight-at-length from fitting a power function. A single model was fit to all samples.

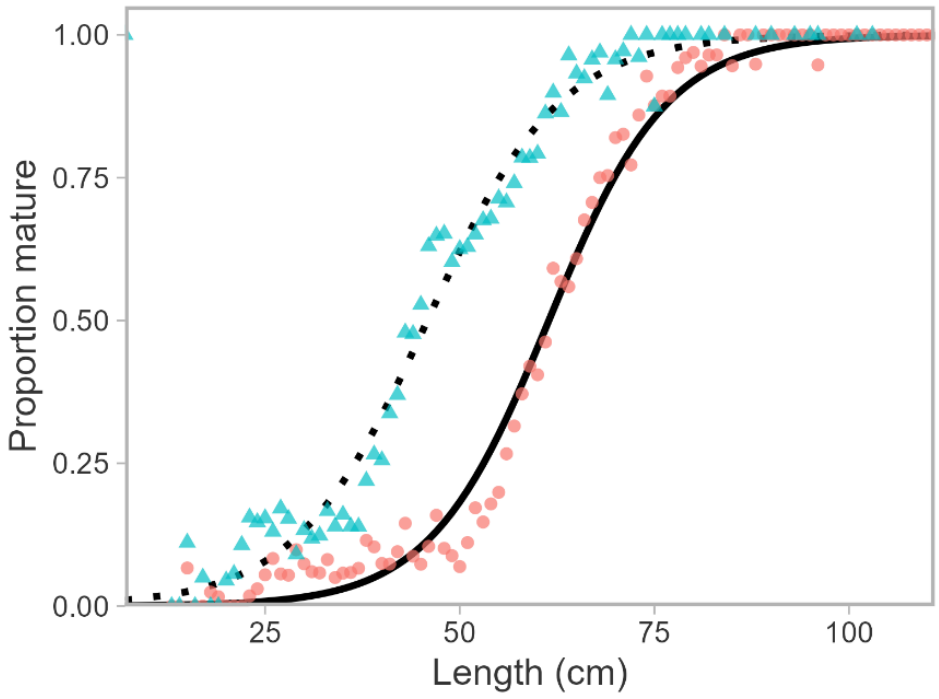

Figure 7. Maturity-at-length (females in red circles, males in blue triangles) of Greenland Halibut from biological samples collected from the trawl survey from 1999-2019. Lines indicate the predicted proportion mature at length (females in solid line, male in dotted line) from a logistic regression.

From these data and parameters, RCM estimated the initial abundance and annual recruitment as deviations from mean recruitment, along with selectivity parameters for the fishery and survey, and fishing mortality by year and fleet. The selectivity function for the trawl gear was

dome-shaped as the length composition did not have large fish that were observed in the gillnet and longline fisheries, with the parameters specifying the extent of the dome estimated in the population model. In the absence of age composition, the model converted the predicted age composition to length composition for the likelihood of the observed length data.

2.2. SPATIAL OPERATING MODEL

Next, the spatial distribution of the stock was characterized in areas of 400–1500 m depth in Divisions 0AB and 1A-D, from 62–73º N and 50–75º W (Figure 8). The bathymetry raster data from the ETOPO1 Ice Surface Global Relief Model was downloaded from the NOAA National Centers for Environmental Information website. The raster data were aggregated into 10 km x 10 km cells (n_s = 2,949 cells) with the depth calculated as the mean value from the smaller constituent cells. While there may be variation in depth especially in steep slope bathymetry, some approximation is needed based on the spatial resolution of the grid cells.

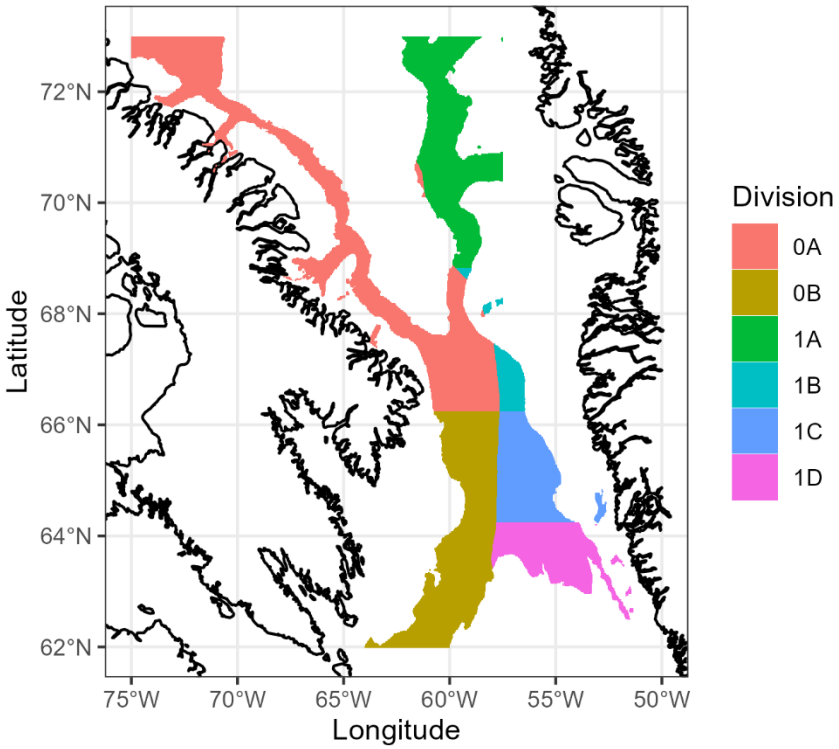

Figure 8. NAFO Divisions for the spatial grid (between 400–1500 m of depth) of the survey simulation.

The spatial distribution was obtained by fitting a spatiotemporal generalized linear mixed model (GLMM) to the RV *Paamiut* survey data (individual trawl sets) in Divisions 0A-South and 1CD. For this analysis, spatial coordinates were available for 1CD only during 2003-2017. The spatiotemporal GLMM was implemented in sdmTMB (see Appendix A for description). For each survey sample, the model predicted catch rate is comprised of fixed effects along with a spatial and spatiotemporal random effect. The fixed effects included an intercept, year, depth, and the square of depth terms, as catch rates were observed to be parabolic with depth (Figure 9).

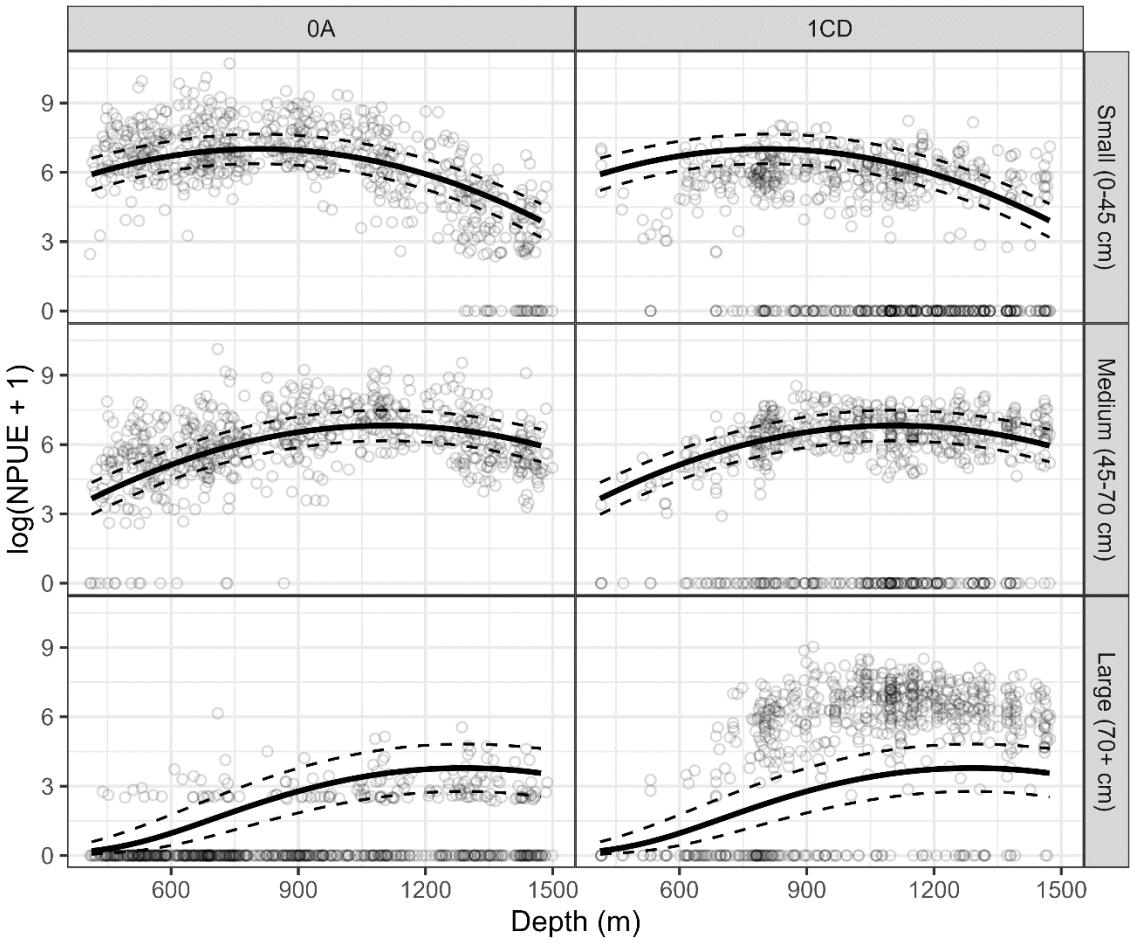

Figure 9. Historical observed NPUE (numbers per sq. km swept area, in open points) by size class (rows) and depth (x-axis). Observed NPUE are shown separately for Divisions 0A-South and 1CD (columns). Solid, bold lines indicate the marginal effect of depth from the spatiotemporal model (averaged across years and excluding all spatial and spatiotemporal effects). The dotted lines denote the 95 percent confidence interval. Marginal effects are identical for both Divisions 0A-South and 1CD. The y-axis is transformed by taking the natural logarithm of NPUE + 1.

The random effects component is intended to capture processes that affect abundance but are not accounted for in the fixed effects. The spatial effects vary in space but are constant through time and the spatiotemporal effects vary in both space and time. These effects are estimated as separate random fields estimated from a mesh of 438 vertices defined across the spatial domain (Figure 10). The spatial and spatiotemporal effects for sample i are obtained by bilinear interpolation based on the UTM coordinates from the sample and the nearest vertices of the spatial mesh.

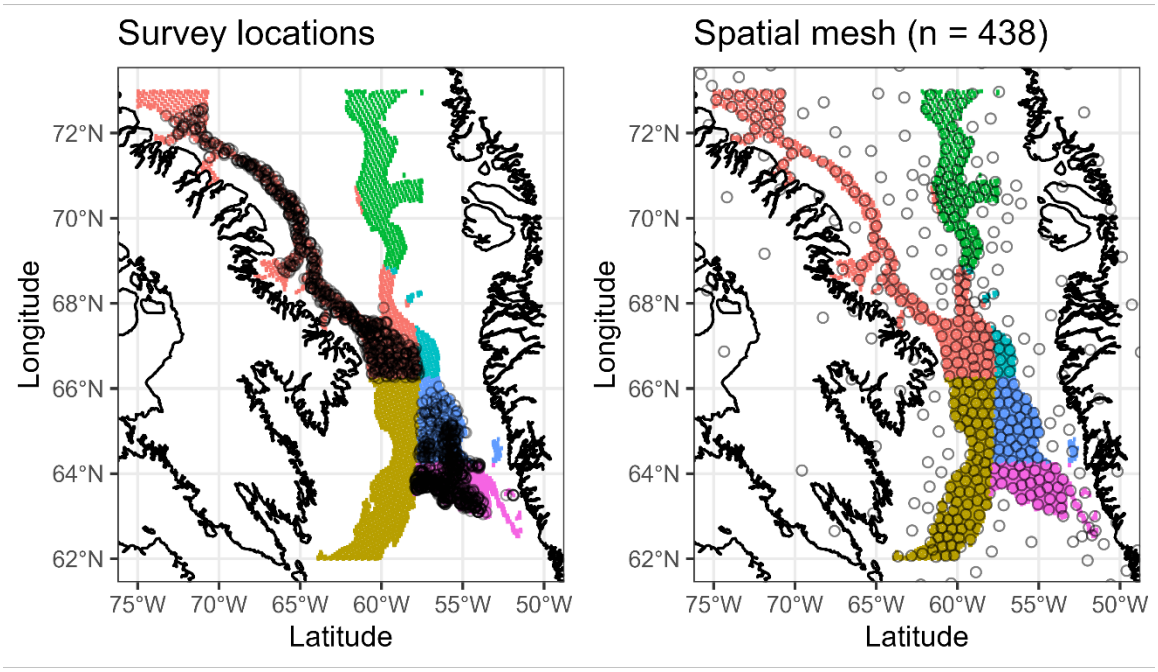

Figure 10. Left: a map of the spatial location of the trawl survey sets in the spatiotemporal model. Right: locations of the 438 points in the spatial mesh where the spatial and spatiotemporal random effects are estimated. The random effects for any other location is obtained by interpolation using the three nearest mesh points. For both, colored regions indicate the NAFO Division following Figure 8.

Based on catch rates between 0A-South and 1CD, it was observed that distribution was not identical for all size classes, with smaller fish somewhat more abundant in 0A-South than in 1CD. Conversely, larger fish are much more abundant in 1CD than in 0A-South (Figures 9, 11). To account for spatial heterogeneity in catch rates by size, catch rates (number caught per square km swept area) for each tow were calculated for three size classes: Small (0–45 cm), Medium (45–70 cm), and Large (70+ cm). This pattern implies an interaction effect between size class and the spatial random field. However, as this interaction could not be modeled in a single GLMM in sdmTMB, separate models were fit for each size class. This implicitly results in interaction effects between size class and all fixed effects and spatial fields, with separate nuisance parameters (dispersion parameter of the likelihood, marginal standard deviation of the random fields, Matérn range that controls the distance at which the random effects are correlated) by size class.

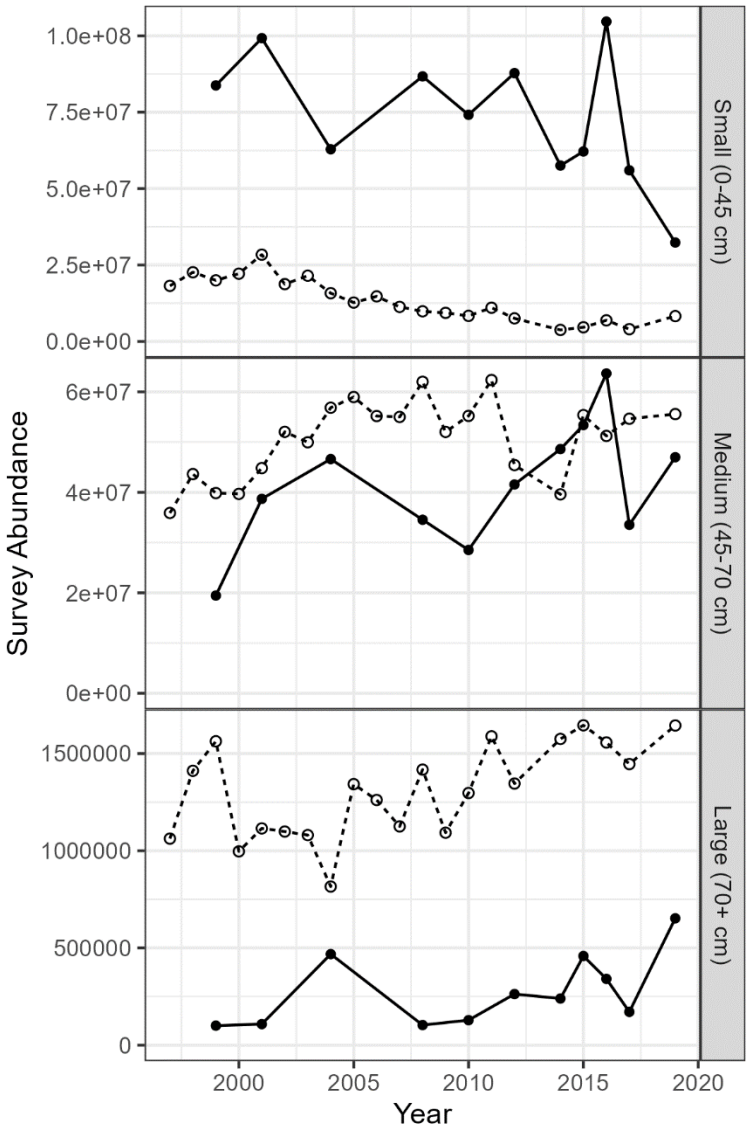

Figure 11. Stratified abundance estimates from the trawl survey by division (filled circles and solid lines for Division 0A-South; open circles and dotted line for Division 1CD) and size class. The y-axis for the three size classes are not identical. This figure includes the 2019 survey with the Helga Maria fishing vessel, the Paamiut research vessel was used in all other years.

From the fitted spatial GLMMs, the population density (numbers per square km) was predicted across the full spatial domain. The distribution $p_{s,t}$ was calculated as

$$p_{s,t,\ell} = \frac{\mu_{s,t,\ell}}{\sum_s \mu_{s,t,\ell}}$$

where $\mu_{s,t,\ell}$ is the predicted density for size class ℓ in cell s. Distribution was calculated for all years of the age-structured model (1987-2020). For years for which there were no survey data (i.e., years not included in the spatiotemporal model), the abundance was predicted based on the intercept, depth, and spatial effects only (excluding the year fixed effect and spatiotemporal random effects).

These distributions were then used to generate the spatial abundance in the spatial operating model, with

$$N_{s,t,a} = N_{t,a} \times p_{s,t,a}$$

Size classes were assigned to ages using the mean length-at-age function, with the Small size class assigned to ages 0–7, Medium to 8-22 years, and Large to 23+ years. Cohort slicing was used because annual age-length keys, specifying the probability of age-at-length, were not available.

These inputs were provided to the SimSurvey package in the `*sim_abundance()*` and `*sim_distribution()*` functions.

2.3. SURVEY SIMULATION

The survey protocol was specified in SimSurvey by identifying the stratum and division of each spatial cell in the domain. In Division 0A-South and 0B, the strata were defined by both area and depth, the latter in increments of 200 m (Table 2). In 1C and 1D, the strata were only defined by depth. For each year, the sampling intensity was set to one sample per 750 sq. km, with a minimum of two samples, per stratum. The swept area was set to 0.08 sq. km, the median value from the historical data. These are input arguments to the `*sim_survey()*` function.

One hundred simulations were run to generate stochastic samples of the random-stratified survey. The simulated catch n was sampled from a binomial distribution

$$n_{x,t,a,i} \sim \mathrm{Bin}\left(N_{s_{x,i},t,a}\,, q_a \frac{A_i}{A_{s_{x,i}}}\right)$$

where $s_{x,i}$ is the spatial cell (randomly chosen within strata) sampled in simulation x and sample i, q_a is the catchability of the survey using the estimated selectivity from the age-structured model, $A_{s_{x,i}}$ = 100 sq. km is the area of the sampled cell and A_i = 0.08 sq. km is the swept area of individual survey tows.

The simulation samples the entire spatial domain. Two indices were generated by adjusting the spatial coverage of the survey (i.e., filtering the SimSurvey output for the desired grid cells). The first index used samples from only Divisions 0A-South and 1CD following the historical coverage of the survey. The second index also included Division 0B, i.e., 0A-South, 0B, and 1CD, as an example of evaluating increased survey coverage into another division. The number of sets per stratum in Divisions 0A, 1C, and 1D were adjusted to match values in Table 2, based on the recent sampling of the actual surveys (Table 1 of Treble 2020; Table 4 of Nogueira and Estevez-Barcia 2020). Once the samples were filtered to match the desired spatial coverage, the stratified totals were calculated for the respective survey.

Table 2. The number of sets simulated in each stratum per year.

Division	Stratum ID	Depth Stratum	Number of Sets
0A	A0	1000	2
0A	A0	1200	2
0A	A0	1400	2
0A	A0	400	2
0A	A0	600	2
0A	A0	800	2
0A	A1	1000	6

Division	Stratum ID	Depth Stratum	Number of Sets
0A	A1	1200	3
0A	A1	1400	2
0A	A1	400	6
0A	A1	600	5
0A	A1	800	6
0A	A2	1000	3
0A	A2	1200	3
0A	A2	1400	3
0A	A2	400	5
0A	A2	600	3
0A	A2	800	3
0A	A3	1000	2
0A	A3	1200	2
0A	A3	1400	2
0A	A3	400	2
0A	A3	600	2
0A	A3	800	2
0A	A4	1000	2
0A	A4	1200	2
0A	A4	1400	2
0A	A4	400	2
0A	A4	600	4
0A	A4	800	2
0A	A5	1000	2
0A	A5	1200	2
0A	A5	1400	2
0A	A5	600	2
0A	A5	800	4
0A	DF	400	2
0B	A1	600	3
0B	B0	400	2
0B	B1	1000	10
0B	B1	1200	7
0B	B1	1400	2
0B	B1	400	22
0B	B1	600	9
0B	B1	800	13
1C	-	1000	2
1C	-	400	2
1C	-	600	10

Division	Stratum ID	Depth Stratum	Number of Sets
1C	-	800	7
1D	-	1000	17
1D	-	1200	11
1D	-	1400	4
1D	-	400	2
1D	-	600	2
1D	-	800	4

2.4. PERFORMANCE MEASURES

To compare the two surveys, four comparisons were made between the simulated indices and the operating model:

1. the relative trends in the index vs. the relative trends in total abundance;
2. the relative trends in the index vs. the relative trends in total vulnerable abundance (accounting for selectivity of the trawl gear);
3. the *r* ratio in the index vs. in total abundance; and,
4. the *r* ratio in the index vs. in total vulnerable abundance.

Fishery total allowable catches (TACs) have recently been set according to a harvest control rule of the form $\mathrm{TAC}_{t+1} = r_t \times \mathrm{TAC}_{\mathrm{t}}$ (Treble and Nogueira 2020), where r is the ratio of index (I_y) means over time:

$$r_t = \frac{\frac{1}{3}\sum_{y=t-2}^{t} I_y}{\frac{1}{4}\sum_{y=t-6}^{t-3} I_y}$$

The values of r_t are compared with the corresponding ratios in the population (replacing I_y with abundance in the above equation).

For each comparison, the relative root mean square error (RMSE) metric was used to gauge how well the index and control rule follow population trends. RMSE was calculated as

$$\text{Relative RMSE} = \sqrt{\frac{\sum_x \sum_t \left(\frac{S_{x,t} - O_{x,t}}{O_{x,t}}\right)}{X\,T}}$$

With the simulated index value S compared to operating model value O over $X = 100$ simulations and T years. The relative RMSE was calculated on the relative trends for the operating model abundance and simulated indices, obtained by dividing annual values by the time series geometric mean.

3. RESULTS

3.1. OPERATING MODEL

In the population model, the selectivity of the trawl was estimated to be dome shaped at the modal age of 13 years, and low selectivity of 0.1 and less for ages older than 30 years (Figure 12). There was variability in the estimated recruitment over time with a peak in 2000, although

there did not appear to be any particular directional trend over time (Figure 13). This appears to be inferred by the relatively flat index over the course of the time series (Figure 14).

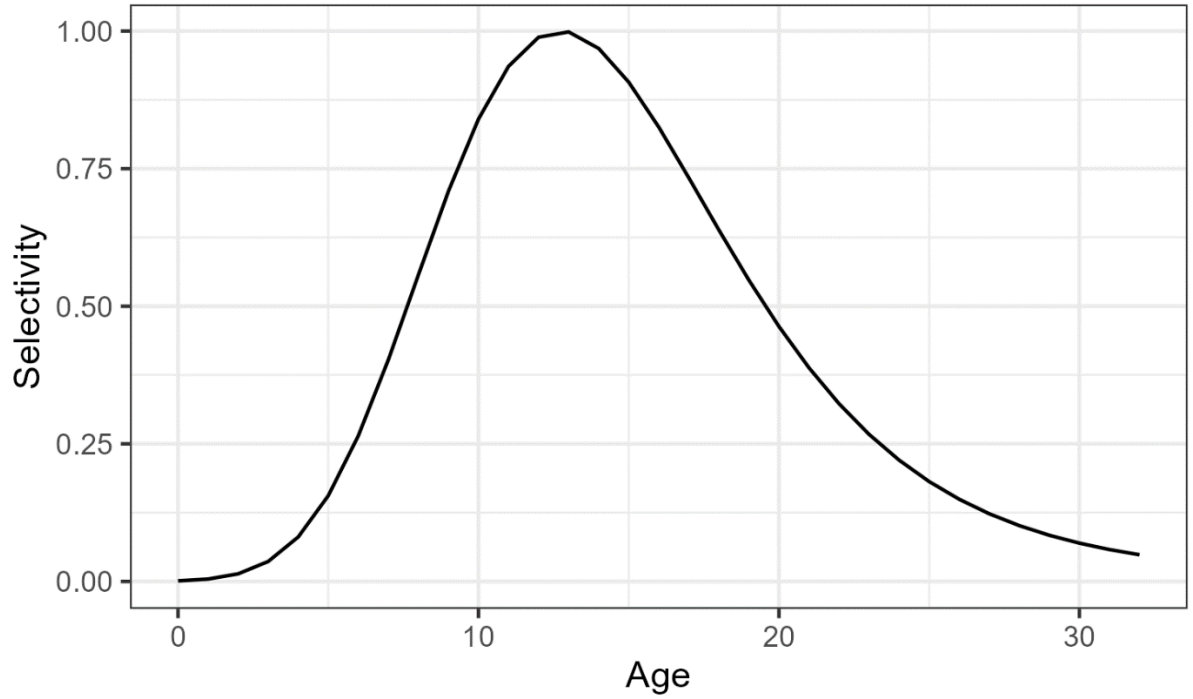

Figure 12. Trawl survey selectivity-at-age estimated in the population model.

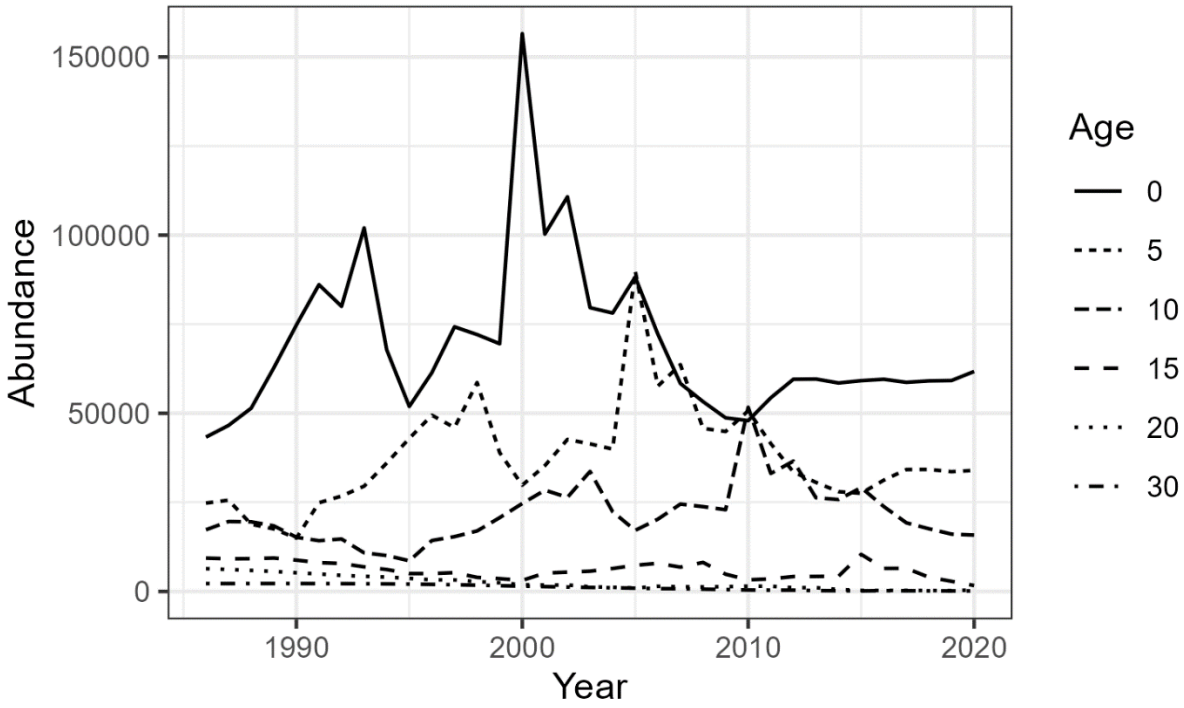

Figure 13. Historical abundance of six age classes from the population model.

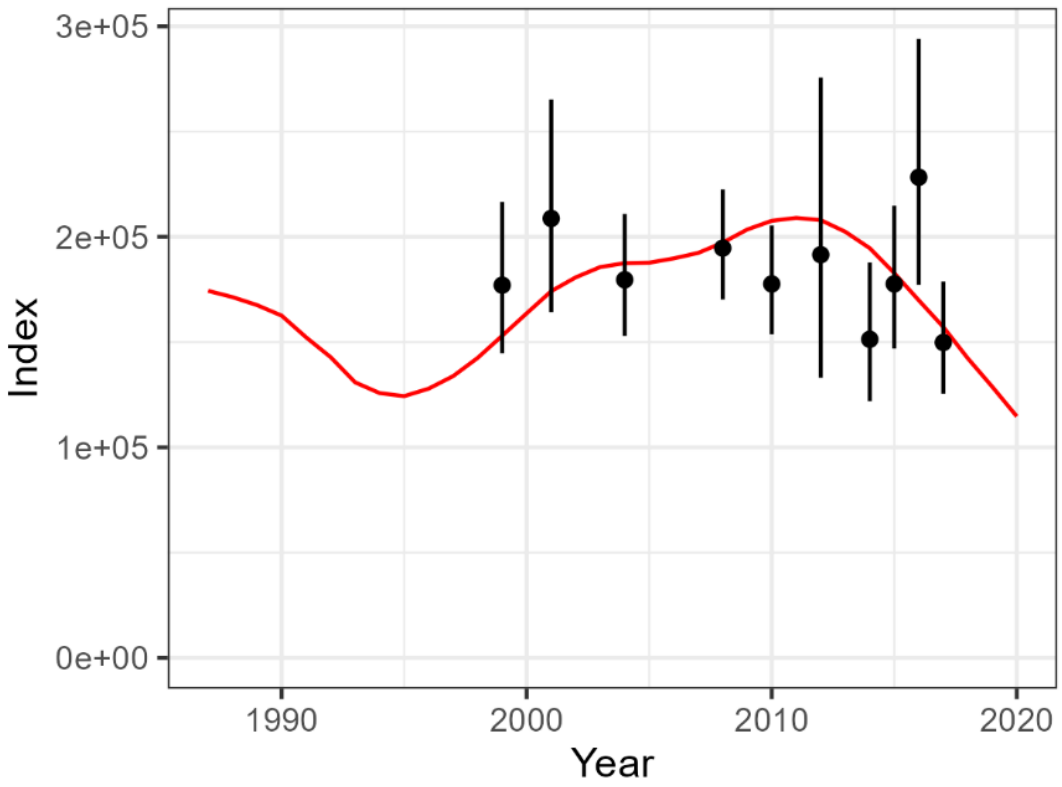

Figure 14. Observed (black points, with error bars indicating the 95% confidence interval) and predicted (red line) trawl survey index in the population model. The 2019 survey was excluded from the Rapid Conditioning (RCM) Model.

Based on the predicted catch rates for the survey samples, the spatiotemporal GLMM captured the size class and spatial interaction effects, with smaller fish predicted to be more abundant in shallow waters and in 0A and less abundant in deep waters and 1CD (Figures 9, 15). This is consistent with Greenland Halibut distribution in other areas (Morgan et al. 2013, Wheeland and Morgan 2020). Conversely, larger fish were abundant in deep waters and 1CD. The depth fixed effects in the GLMM were significant at the 95% level (Table 2). This spatial effect is also apparent in the estimated spatial random field with a positive effect for small fish in 0A and for large fish in 1CD (Figure 15). The model interpolates the spatial random field in Divisions 0B, 0A-North, and 1AB as there were no survey data informing abundance in these regions.

The spatiotemporal field estimated year-specific deviations in abundance, particularly in the medium and large size classes (Figure 16). Notably, there was no spatiotemporal variation estimated for the small size class. The estimated standard deviation of the spatiotemporal random field was low (SD = 0.02, Table 3), presumably because the fixed effects and spatial field can account for most of the variation in catch rates.

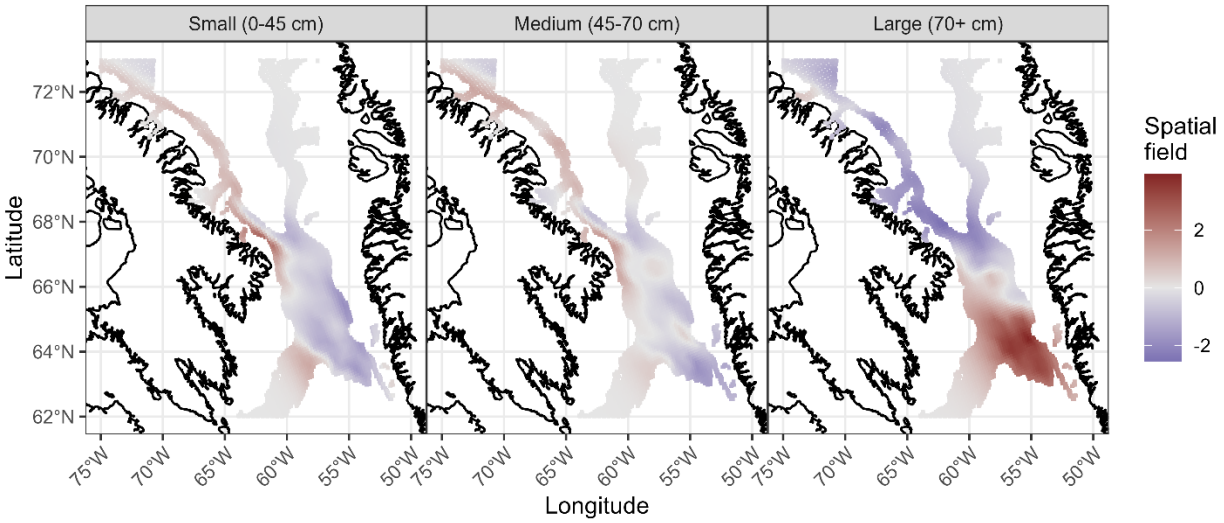

Figure 15. The estimated spatial random field (in log space) in the GLMM by size class. Higher values of the random field predict higher abundance in the corresponding spatial cell.

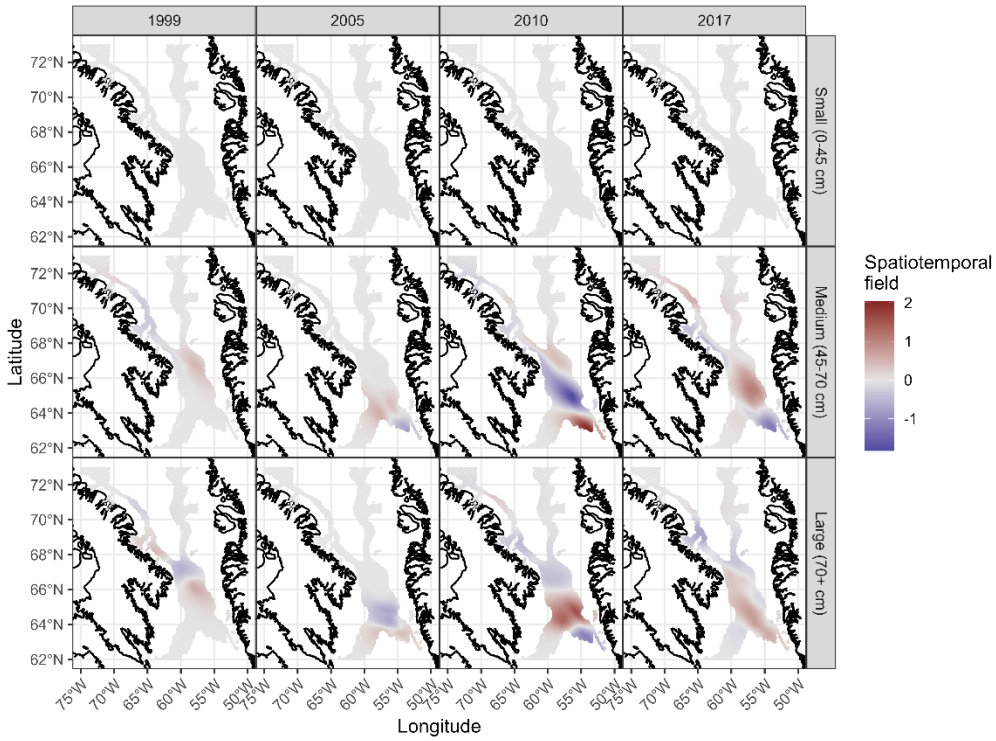

Figure 16. The estimated spatiotemporal random field (in log space) in the GLMM by size class for four years. Higher values of the random field predict higher abundance in the corresponding year and cell.

Figure 17 shows the density predicted over the spatial domain for a subset of 4 different years from the model. Overall, there was no notable inter-annual changes in the predicted spatial density, supported by the lack of significance in the year effects in the GLMM (Table 3). Overall,

the model informs the spatial distribution through the depth and area effects and their interactions by size class.

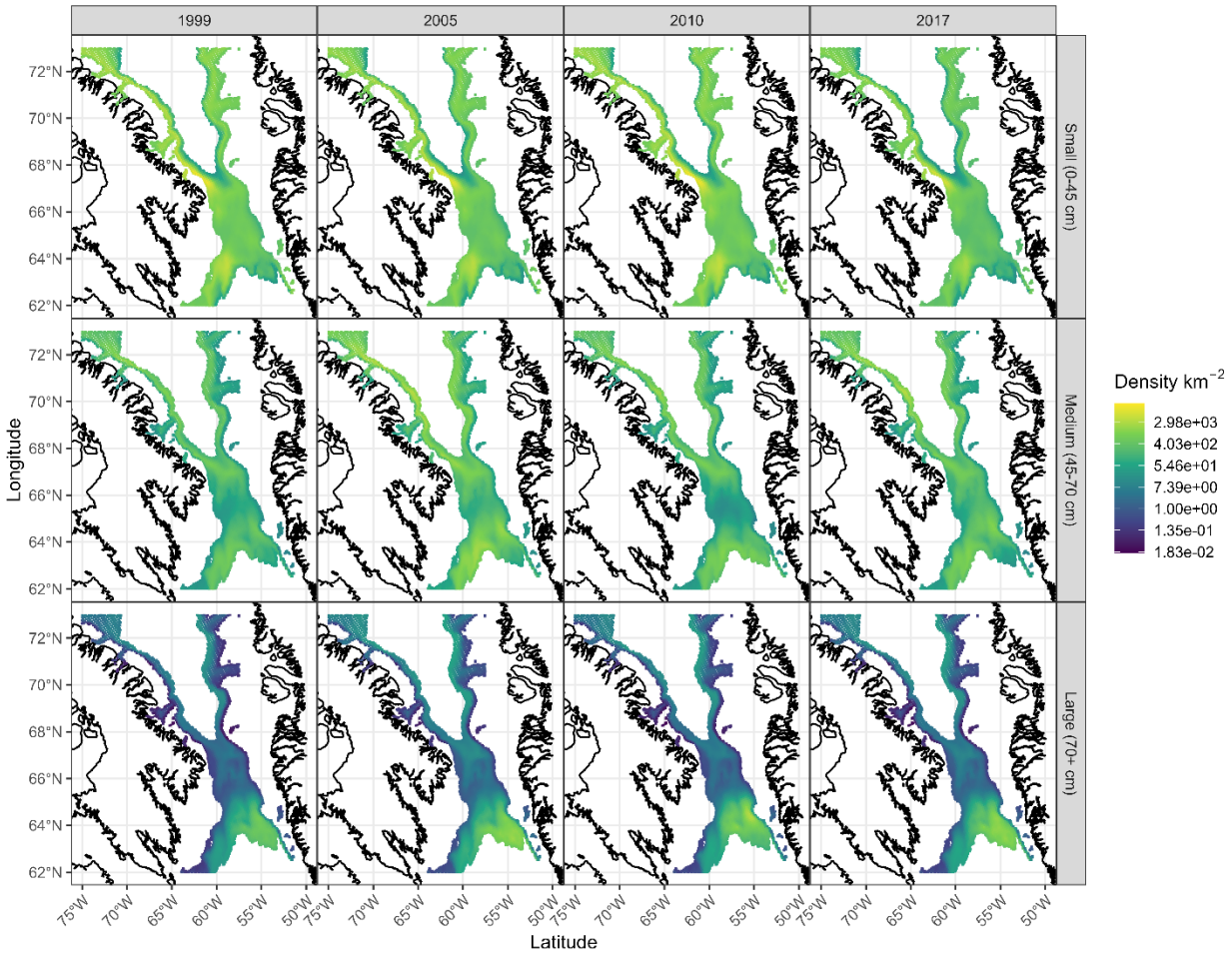

Figure 17. Estimated stock density (numbers per sq. km) over spatial area by size class for four historical years (1999, 2005, 2010, and 2017).

Table 3. GLMM estimates of fixed effects (in log space) and nuisance parameters. Separate models were fit for survey catch rates for each size class (Small, Medium, Large). Asterisks indicate coefficients that were significant at the 95% level (null hypothesis of zero). `depth_z` is the tow depth converted to a Z-score and `depth_z2` is the square of `depth_z`. Significance is not reported for nuisance parameters.

Term	Estimate		
	Small	Medium	Large
(Intercept)	7.16*	6.12*	2.62*
year2001	-0.06	0.81	-0.19
year2003	-0.06	1.26*	0.02
year2004	-0.35	1.04*	0.79
year2005	-0.22	0.79	0.87
year2006	-0.52	0.90	0.08
year2007	0.15	0.94	0.15

Term	Estimate		
	Small	Medium	Large
year2008	-0.21	0.23	0.32
year2009	-0.53	0.75	1.05
year2010	0.01	0.17	0.43
year2011	-0.33	0.87	0.83
year2012	-0.44	0.61	0.46
year2013	-0.13	0.59	-0.03
year2014	-0.87*	0.44	0.62
year2015	-0.58*	1.04*	0.87
year2016	-0.41	0.73	0.56
year2017	-0.57*	0.28	0.47
depth_z	-0.72*	0.46*	1.20*
depth_z2	-0.58*	-0.54*	-0.56*
Matérn range (km)	1463	1782	2125
Dispersion (ϕ)	0.56	0.81	0.83
Spatial SD	1.12	0.98	1.38
Spatiotemporal SD	0.02	0.71	0.92

3.2. SURVEY SIMULATION

Figure 18 shows three simulated indices with index values generated for all years 1987-2020 (years in the population model). Performance measures were calculated using simulated values from all years. A subset of values corresponding to years in which the survey was actually run can be used as a “Turing test” to visually evaluate if the simulation reasonably mimics the true index (Figure 19).

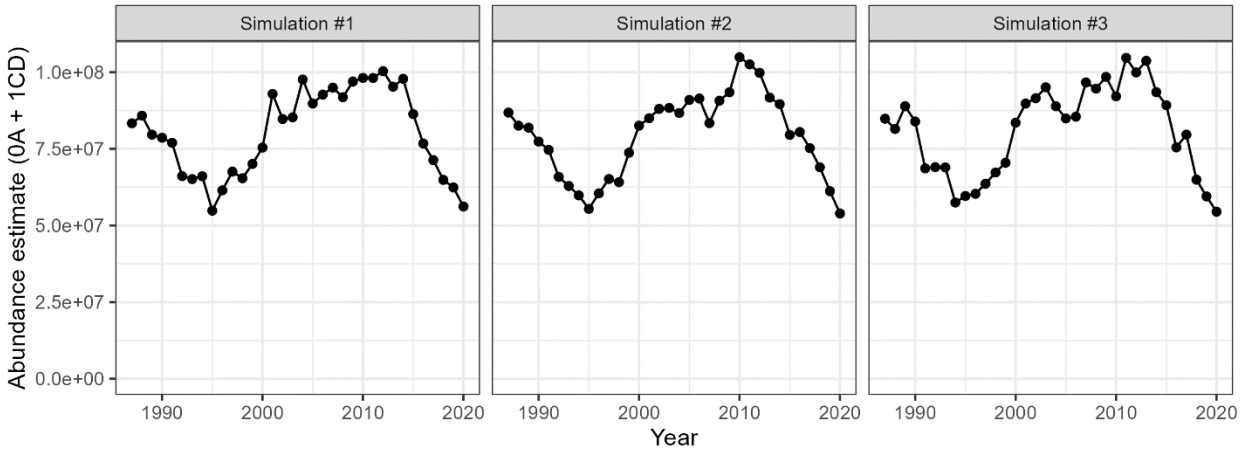

Figure 18. Example index generated from three simulations in SimSurvey. The index was generated from sampling Divisions 0A-South and 1CD using the recent sampling intensity (by depth-area stratum).

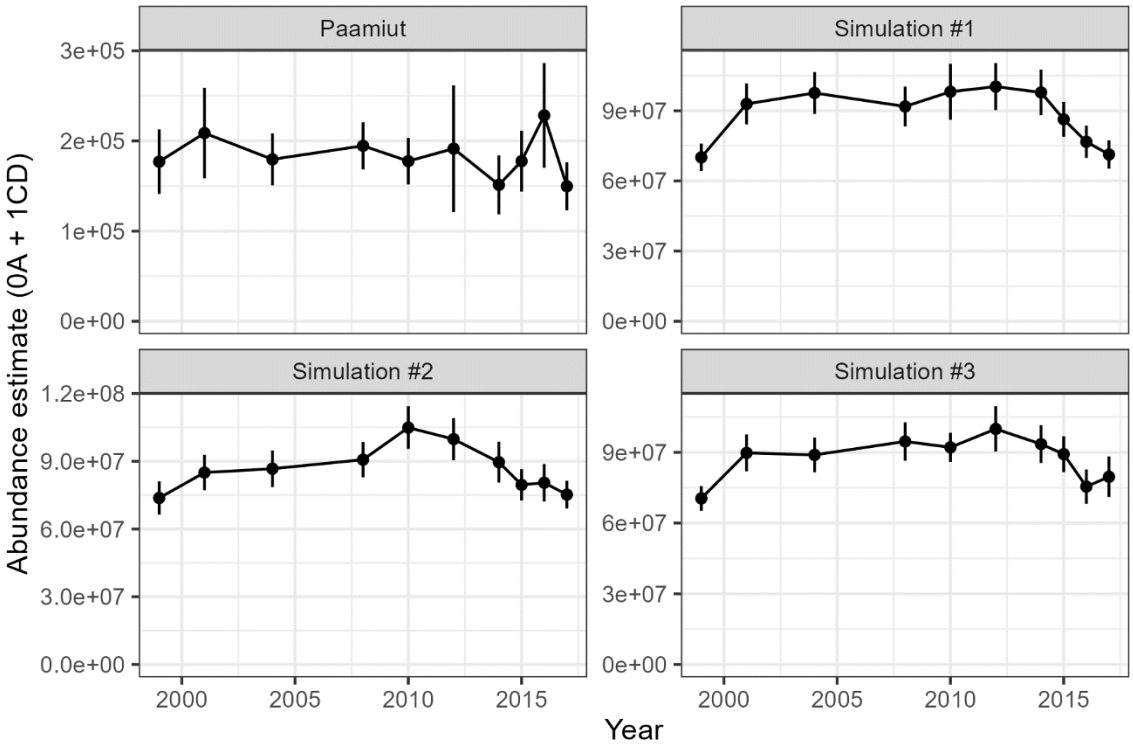

Figure 19. Example index with 95% confidence intervals generated from three simulations, along with the historical stratified estimates from the RV Paamiut *survey. A subset of values is plotted with years in which the actual survey took place. Differences in scaling between the simulated indices and historical index could be attributed to specification of catchability-at-age in SimSurvey.*

When comparing the index to total stock abundance, the precision of the simulated index was slightly more precise when 0B was sampled than when omitted from the survey (Figure 20). However, the reduction in relative RMSE was minor (0.18 vs. 0.175). More importantly, the index trend is out of phase with the operating model, presumably due to selectivity and the delay between recruitment of young animals to the population versus when they are vulnerable to the survey. When compared to vulnerable abundance, the relative RMSE is smaller, with greater reductions in error to 0.036 from 0.046 when Division 0B is included in the survey (Figure 21).

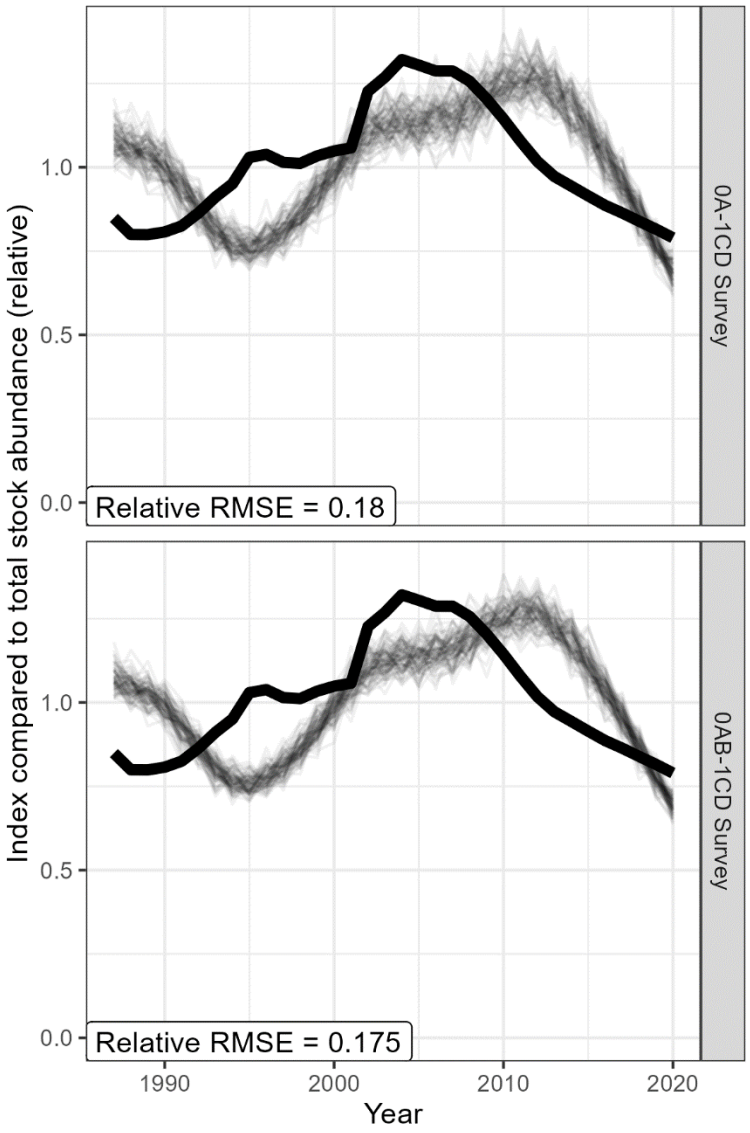

Figure 20. Comparison of total abundance in the operating model (thick black line) to two simulated surveys. Thin lines represent 100 simulated indices with the relative root-mean-square error (RMSE) in the lower left corner. There is little improvement in the relative RMSE with additional spatial coverage because the simulated index is out of phase with the population abundance due to dome selectivity.

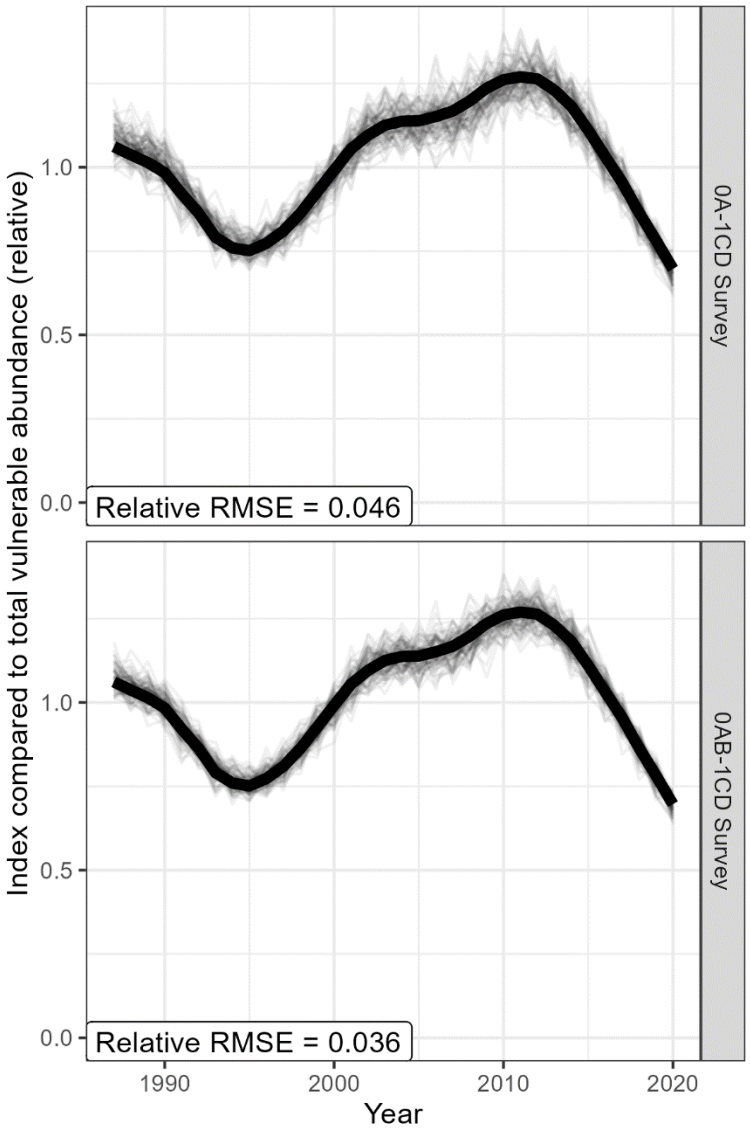

Figure 21. Comparison of total vulnerable abundance in the operating model (thick black line) to two simulated surveys. Thin lines represent 100 simulated indices with the relative root-mean-square error (RMSE) in the lower left corner. It is expected that relative RMSE will reduce with additional spatial coverage; more information on survey constraints are needed to identify the cost and benefits of alternative sampling protocols.

Lower RMSE values in r_t indicated that the ratio was somewhat more precise than individual index values (Figure 22-23). Trends in r_t remain out of phase with trends in total stock abundance and there was little reduction in error with increasing spatial coverage. Incorporating Division 0B in the survey reduces the error of r_t in tracking vulnerable abundance by 21% (reduction in relative RMSE from 0.034 to 0.027).

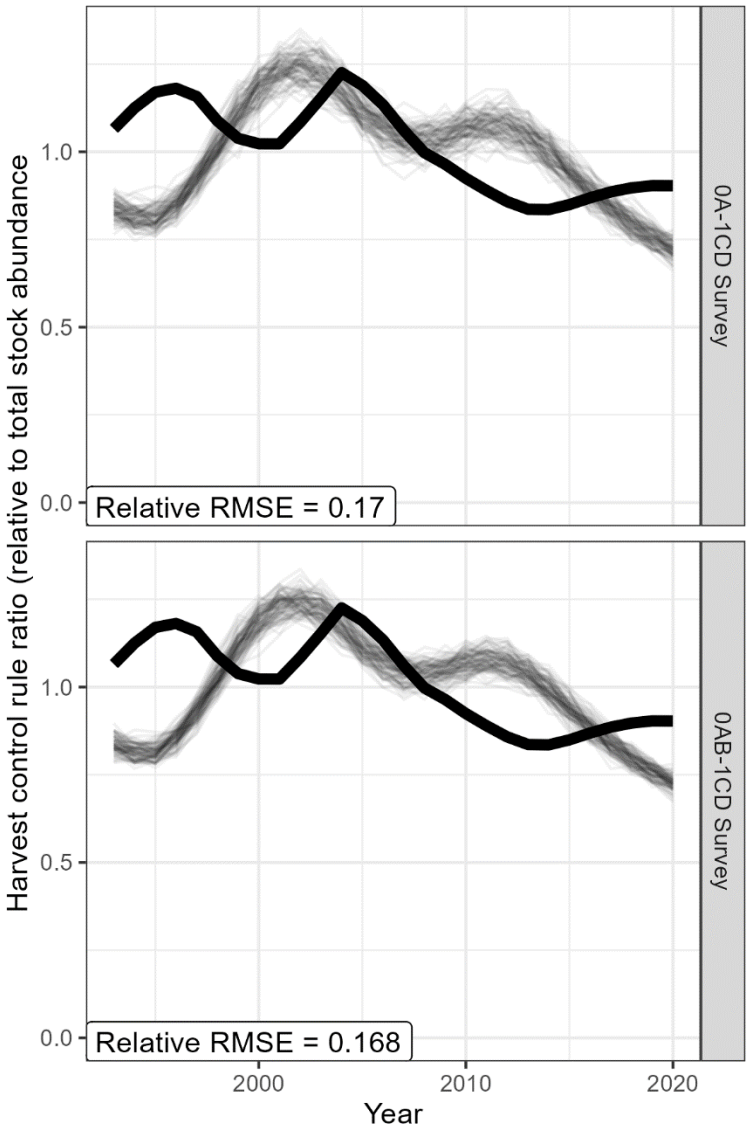

Figure 22. Comparison of index ratio r from two simulated surveys to the ratio of total abundance means in the corresponding year from the operating model (thick black line). Thin lines represent 100 simulated indices with the relative root-mean-square error (RMSE) in the lower left corner. There is little improvement in the relative RMSE with additional spatial coverage because the simulated index is out of phase with the population abundance due to dome selectivity.

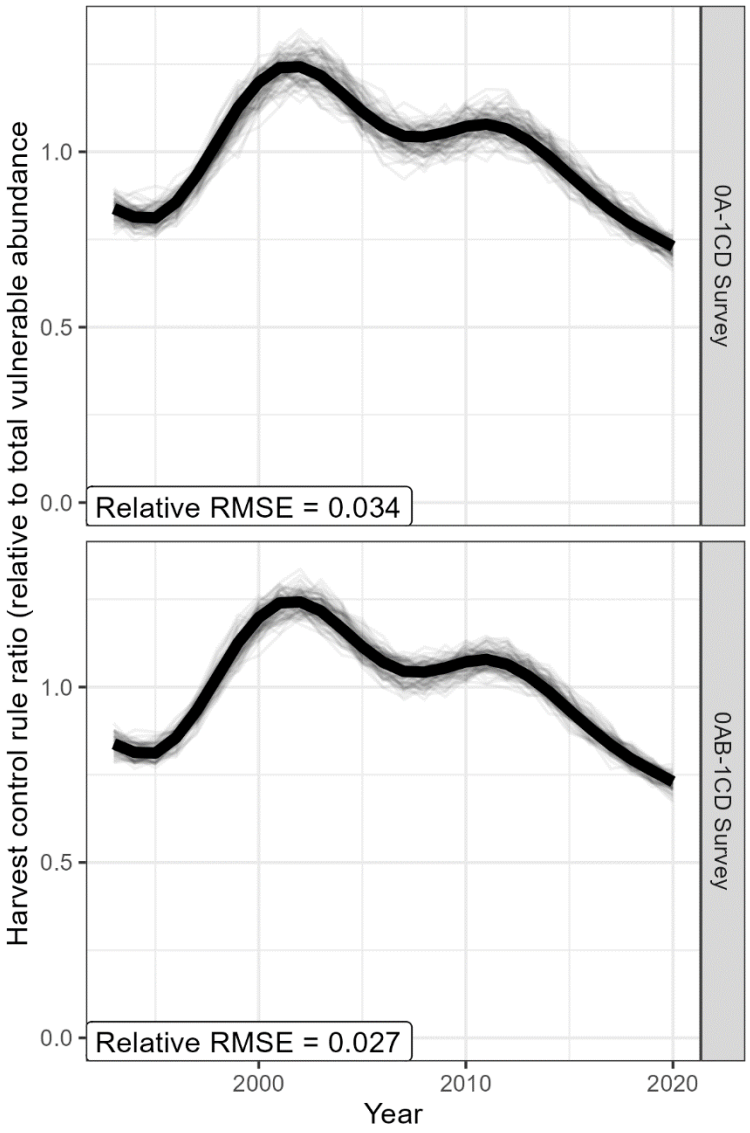

Figure 23. Comparison of index ratio r from two simulated surveys to the ratio of total vulnerable abundance means in the corresponding year from the operating model (thick black line). Thin lines represent 100 simulated indices with the relative root-mean-square error (RMSE) in the lower left corner. It is expected that relative RMSE will reduce with additional spatial coverage; more information on survey constraints are needed to identify the cost and benefits of alternative sampling protocols.

4. DISCUSSION

Factors such as ship availability, funding for survey operations, and access to ice-free waters can constrain the ability to sample the population. In turn, they may affect the spatial coverage and sampling intensity of the survey. With these logistical constraints, the recent change to a new survey vessel provides an opportunity to re-evaluate survey design and implement new protocols if desired. Given the high cost and complexity of implementing a survey, the ability to evaluate a survey design and its effect on estimating population trends is highly desirable and can be used to justify the change to a more effective sampling protocol.

This paper demonstrated the steps in setting up a simulation for evaluating survey design using Greenland Halibut as a case study. The spatial operating model was developed from two components: (1) a population model with estimates of historical stock abundance-at-age (conditioned on the data), and (2) a spatial distribution model conditioned on historical survey tows. Sampling protocols can then be implemented to simulate a survey to evaluate the performance and benefit of alternative survey designs that differ in spatial boundaries of sampling strata, sampling intensity, and spatial coverage.

4.1. OPERATING MODEL SPECIFICATION AND EVALUATION OF SURVEY DESIGNS

Conditioning the population model with catches from 0AB and 1A-D allowed for estimates of total abundance in a spatial region larger than those sampled by the offshore survey in 0A-South and 1CD. Some assumption of the distribution in regions not sampled by the actual survey was needed. The species distribution model used here allows one to interpolate in missing regions.

Use of the population and distribution models would allow one to evaluate robustness of a survey design with respect to abundance trends and spatial distribution. Alternative conditioning of population models can evaluate robustness with respect to productivity and stock depletion. The distribution model estimates the historical distribution of the stock. In this way, the simulation evaluates the retrospective performance of any survey design. This setup evaluates what sampling plan would generate the best index, given historical abundance and distribution. In the case study here, separate spatiotemporal models were fit to account for variation in spatial distribution by size. However, the data did not show much inter-annual variation in spatial distribution and patchiness, factors that can be expected to have notable impacts on a survey with respect to spatial coverage and sampling intensity.

The SimSurvey operating model can be expanded to include historical survey data that is available for several years from 0A-North, 0B and 1AB as well as a longer time series from nearshore areas of Divisions 1A-F (Hedges and Raffoul 2023). These data could be incorporated into the spatiotemporal model to inform the spatial distribution of Greenland Halibut. It is known that that there is spatial structure of Greenland Halibut in Subarea 0+1 between the offshore and nearshore component. Operating models can be updated to evaluate survey performance with consideration to these dynamics.

More broadly, survey design should take into account the spatial dynamics of the sampled species. For example, the best survey design for a sessile species may not be the best for a mobile species. A mobile species may exhibit high variability in patchiness over space and time, in which case, broad survey coverage may be desirable. Functions in SimSurvey such as `*sim_ays_covar()*` allows for simulation of spatial patchiness over time in lieu of estimating distribution from sdmTMB. Mechanistically, distribution matrices derived from movement matrices would explicitly account for the ‘viscosity’ of a population (Carruthers et al. 2015). Tagging models can inform movement between offshore regions. Such movement matrices typically describe movement among large-scale spatial units, e.g., at the division level, with additional specification needed for individual fine-scale spatial cells within division.

This paper evaluated two alternative survey designs using only root mean square error in the objective function to evaluate performance. Unsurprisingly, performance should always improve with additional sampling. For effective use of SimSurvey to inform survey planning, the logistical constraints of the survey should first be identified and subsequently operationalized within the simulation. For example, the number of available sea days in the survey season sets a constraint on the number of possible sets per year. Within those constraints, one can then

compare the benefits of various options, e.g., more sets within a concentrated area or fewer sets in a broader area. If budgetary costs are a concern, then a cost-benefit tradeoff can be incorporated into the analysis to justify expenditures of additional sampling.

4.2. DISCONTINUITY IN SURVEY PROTOCOLS

The SimSurvey simulation model was developed to address the opportunity for potential revisions to survey design that arose from the change to a new vessel. However, there is a break in the index series that cannot be bridged by a calibration factor typically derived from gear comparative studies. We propose two methods to address this index gap.

First, the spatiotemporal model can be modified to estimate a calibration factor without comparative towing. By constraining the year effect explicitly (intercepts as a random walk), or implicitly (year-over-year random walk in the spatiotemporal random effects), abundance may not be confounded with catchability. In essence, the random walk models the population process (changes in abundance) as informed by the older index series. The statistical properties of the random walk, i.e., how much the abundance can change, informs the predicted abundance after the end of the old index series (including the intervening years between surveys). Deviations in catch rates observed in the new survey from the predicted abundance inform catchability differences. A demonstration of the model-based calibration technique on the 2019 survey that used the *Helga Maria* fishing vessel is provided in Appendix B, which estimated lower catchability in several depth strata, as supported by field observations (Nogueira and Treble 2020). Further analysis by simulation is desirable. It is expected that life history and the length of the survey gap affects performance of the model-based calibration approach. For example, this technique may perform better for long lived species like Greenland Halibut compared to short-lived species which experience rapid turnover in the population. Validation with surveys that have comparative towing (e.g., Miller et al. 2010) will also provide support for using spatiotemporal modeling to estimate calibration factors.

Second, SimSurvey can be used to create a retrospective index from the new survey series to compare with the old series (Figure 24). With population models conditioned on historical data, simulated indices of abundance are generated from the estimated historical abundance. If the old index series was used to condition the operating model, then the simulated indices should resemble the historical index generated from the old vessel. A second set of indices can be simulation using the properties of the new index, i.e., selectivity and catchability estimates of the new vessel. If the two series differ only in magnitude, if at all, then a calibration factor could be derived, for example, using the ratio in the time series geometric mean, to join the two series. If trends appear to differ between the two series within the same historical period, then there is less support to stitch the two indices. Selectivity and catchability estimates can be obtained from either the model-based calibration in the spatiotemporal model or from an age-structured model that fits to the new index series. The extent to which the two series differ in magnitude is expected to be explained by catchability, while selectivity differences may affect trends.

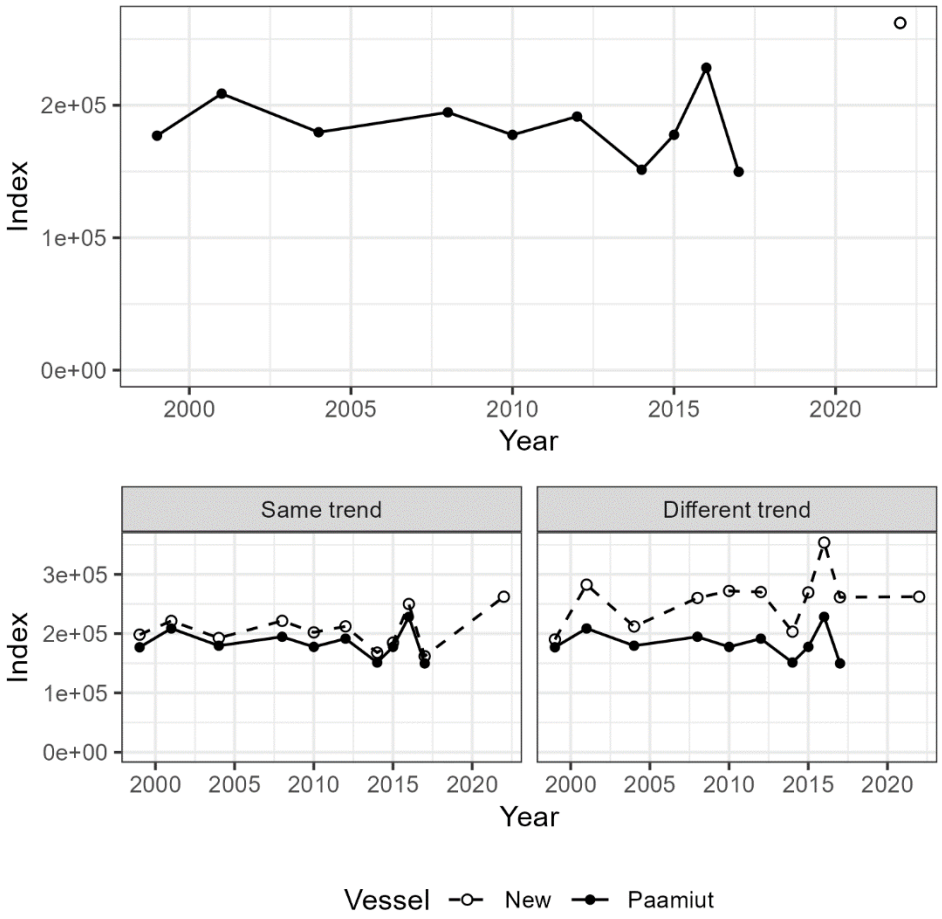

Figure 24. Schematic of the proposed retrospective calibration with SimSurvey. Top: a historical index (Paamiut research vessel) until 2017 (black lines) followed by a hypothetical new index series starting in 2022. Bottom: SimSurvey can be used to simulate a retrospective index series, i.e., the hypothetical index if the new survey vessel could have sampled the population in the past. A calibration factor can be identified if the two indices are similar in trend (bottom left panel). Otherwise, there is less support for stitching the two index series if the two differ in trend (bottom right panel).

4.3. EVALUATION OF HARVEST STRATEGIES

While this study used relative RMSE to evaluate whether an index tracks historical population trends, management strategy evaluation can answer the broader question of whether survey design impacts broader management objectives. It is critical to distinguish between scientific objectives (e.g., accuracy and precision of a survey or stock assessment) and management objectives (e.g., fishery catch and risk of falling below the limit reference point). For example, it is possible that a much more extensive and expensive survey may provide a substantially more precise index of abundance but does not significantly improve a harvest strategy.

Previous catch advice for the stock has been provided from an index-based harvest control rule. With respect to the harvest control rule, two considerations moving forward with the new survey vessel would be (1) how to identify if an alternative survey design supports management, and (2) how to provide catch advice in the interim time period when the control rule would use index values from two separate vessels. We envision that SimSurvey can be incorporated into any management strategy evaluation software such as openMSE for testing in closed-loop

projections. The simulated stock abundance and distribution information can be provided to SimSurvey to generate and update an index of abundance following a specified sampling design at any time step during the projection. This index can be used to specify catch advice in a management procedure that is sequentially implemented in the operating model over time. In this way, error in the index is an emergent property of the stock dynamics and sampling design rather than a pre-specified parameter such as a hyperstability or observation error term. This allows one to evaluate implementation of survey design with respect to management outcomes such as the risk of breaching limit reference points and catch objectives.

Under stationary productivity, the sampling error of the index may be derived from the stratification plan and set density to inform management strategy evaluation simulations. A spatially explicit setup of the operating model with the SimSurvey component would be valuable to evaluate the index when there are systematic changes in stock distribution. Extending this work to undertake closed-loop projections in an MSE framework also allows for robustness testing of management strategies (combined survey design, assessment and harvest control rules) under changing ocean conditions, for example systematic spatial shifts due to climate change. Scenarios can be developed where bottom temperature changes distribution of Greenland Halibut from the historical baseline, which would affect the availability of the stock to the survey. In doing so, it may be possible to establish dynamic survey rules or harvest control rules that are ‘climate ready’. Alternatively, scenarios for climate or ecosystem impacts could be used to develop early warning indicators of important changes in population dynamics.

Finally, construction of operating models provides access to the calculation and comparison of alternative reference point systems in order to identify those that are most appropriate, for example, definitions of a biological limit reference point that is responsive to sustainable/unsustainable fishing or has a low likelihood of being exceeded even with low catches.

5. ACKNOWLEDGEMENTS

The authors wish to thank Margaret Treble and Kevin Hedges from Fisheries and Oceans Canada, and Adriana Nogueira from the Greenland Institute of Natural Resources for providing Greenland Halibut data and context from surveys conducted in Northwest Atlantic Fisheries Organization Divisions 0AB and 1CD. We also acknowledge the participants of the Canadian Science Advisory Secretariat peer review meeting, whose comments substantially improved this research document. Discussions with Sean Anderson on sdmTMB and Paul Regular on SimSurvey were very helpful for this analysis.

6. REFERENCES CITED

Anderson, S.C., Ward, E.J., Barnet, L.A.K., and English, P.A. 2021. sdmTMB: Spatiotemporal Species Distribution GLMMs with 'TMB'. R package version 0.0.21.9005.

Carruthers, T.R., Walter, J.F., McAllister, M.K., and Bryan, M.D. 2015. Modelling age-dependent movement: an application to red and gag groupers in the Gulf of Mexico. Canadian Journal of Fisheries and Aquatic Science 72: 1159–1176.

Hedges, K.J., and Raffoul, D. 2023. Summary of factors that affect survey and fishing catchability and data available regarding the NAFO Subarea 0+1 (offshore) Greenland Halibut (*Reinhardtius hippoglossoides*) stock and fishery. DFO Can. Sci. Advis. Sec. Res. Doc. 2023/037. iv + 11 p.

Hordyk, A., Huynh, Q., and Carruthers, T. 2021. openMSE: Open Source Software for Management Strategy Evaluation. R package version 1.0.0.

Huynh, Q., Carruthers, T., and Hordyk, A. 2021. SAMtool: Stock Assessment Methods Toolkit. R package version 1.2.5.

Miller, T.J., Das, C., Politis, P.J., Miller, A.S., Lucey, S.M., Legault, C.M., Brown, R.W., Rago, P.J. (eds). 2010. Estimation of Albatross IV to Henry B. Bigelow calibration factors. Northeast Fish Sci Cent Ref Doc. 10-05. 233 p. Available from: National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026, or online at

Morgan, M.J., Garabana, D., Rideout, R.M., Román, E., Pérez-Rodríguez, A., and Saborido-Rey, F. 2013. Changes in distribution of Greenland halibut in a varying environment. ICES Journal of Marine Science 70: 352–361. doi:10.1093/icesjms/fss179

Morgan, M.J., Regular, P.M., and Ings, D.W. 2019. Greenland halibut (*Reinhardtius hippoglossoides*) in NAFO Subarea 2 and Divisions 3KLMNO: stock trends based on annual Canadian research vessel survey results and an update of the SAM style model. NAFO SCR 19/036. Serial No. N6954. 33 p.

Nogueira, A., and Estevez-Barcia, D. 2020. Results for Greenland halibut survey in NAFO Divisions 1C-1D for the period 1997-2017, and 2019. NAFO SCR Doc. 20/012. Serial No. N7056. 41 p.

Nogueira, A., and Treble, M.A. 2020. Comparison of vessels used and survey timing for the 1CD and 0A-South deep-water surveys and the 1A-F west Greenland shelf surveys. NAFO SCR Doc. 20/015. Serial No. N7060. 44 p.

Regular, P.M., Robertson, G.J., Lewis, K.P., Babyn, J., Healey, B., and Mowbray, F. 2020. SimSurvey: An R package for comparing the design and analysis of surveys by simulating spatially correlated populations. PLoS ONE 15: e0232822.

Then, A.Y., Hoenig, J.M., Hall, N.G., and Hewitt, D.A. 2015. Evaluating the predictive performance of empirical estimators of natural mortality rate using information on over 200 fish species. ICES J. Mar. Sci. 72: 82–92.

Thorson, J.T. 2019. Guidance for decisions using the Vector Autoregressive Spatio-Temporal (VAST) package in stock, ecosystem, habitat and climate assessments. Fisheries Research 210: 143–161.

Treble, M.A. 2020. Report on Greenland halibut caught during the 2019 trawl survey in Division 0A. NAFO SCR Doc. 20/007REV. Serial No. N7051. 27 p.

Treble, M.A., and Nogueira, A. 2020. Assessment of the Greenland Halibut Stock Component in NAFO Subarea 0 + 1 (Offshore). NAFO SCR Doc. 20/038. Serial No. N7086. 31 p.

Wheeland, L.J., and Morgan, M.J. 2020. Age-specific shifts in Greenland halibut (*Reinhardtius hippoglossoides*) distribution in response to changing ocean climate. ICES Journal of Marine Science 77: 230–240. doi:10.1093/icesjms/fsz152

APPENDIX A: DESCRIPTION OF THE SPATIOTEMPORAL MODEL

The spatiotemporal GLMM implemented here is of the form:

$$\log \mu_i = X_{i,j}\beta_j + O_i + \omega_i + \varepsilon_i$$

where μ_i is the predicted abundance from sample i, $X_{i,j}$ is the design matrix of covariates, β_j is the vector of fixed effects, O_i is the offset term (with a fixed coefficient of 1), ω_i is the spatial effect, and ε_i is the spatiotemporal effect. The fixed effects used here were the intercept, year, depth, and the square of depth. The offset term was the log of the swept area. The natural logarithm is the link function to map observed values to their linear predictors.

In the GLMM, a spatial mesh of n_m = 438 points is defined across the spatial domain. The spatial effects ω_m estimated at the vertices of the mesh is parameterized as a Gaussian random field (multivariate normal distribution):

$$\omega_m \sim \mathrm{MVN}(0, \Sigma_\omega)$$

Similarly, the spatiotemporal random field $\varepsilon_{m,t}$, independent among years t, is also a Gaussian random field

$$\varepsilon_{m,t} \sim \mathrm{MVN}(0, \Sigma_\varepsilon)$$

where Σ is the corresponding n_m x n_m covariance matrix. For each pair of points in the spatial mesh, the correlation is defined by a Matérn function and decays as distance between points increases. The Matérn function is defined by the range parameter, the distance at which spatial points are effectively independent, and is shared between both Σ_ω and Σ_ε. The covariance is scaled from the correlation by the marginal standard deviation, with separate parameters estimated for the two fields.

The spatial and spatiotemporal effects for sample i are interpolated from the corresponding random field based on the UTM coordinates of the sample and the vertices of the spatial mesh.

The marginal likelihood of the model is

$$y_i \sim \mathrm{NegBin}(\mu_i, \phi)$$

where y_i is the observed abundance from sample i, and ϕ is a dispersion parameter such that the variance of $y_i = \mu_i + \mu_i^2/\phi$. The objective function integrates across the likelihood of the random effects.

APPENDIX B: SPATIOTEMPORAL MODEL CALIBRATION OF SURVEY CATCHABILITY

To join time series of survey data where catchability has changed over time, typical index standardization procedures rely on comparative tow studies in order to estimate a calibration factor and link an index from two different gears (otherwise, the year effect is confounded with catchability changes). Autoregressive spatiotemporal models provide an opportunity to explore calibration of catchability in the absence of comparative tow studies.

This approach is possible when a process controlling population abundance is constrained, i.e., the spatiotemporal random field, through a random walk over individual time steps. Spatiotemporal fields estimate time-area effects that describe inter-annual variability in stock distribution. For example, a hotspot that persists only in part of the time series would be separate from the time-invariant spatial field. The variance of the random walk should in theory be informed by biological considerations. A sedentary, long-lived species would be expected to have a lower variance (slower changes over time) in the spatiotemporal field compared to a short-lived, mobile species.

In the context of a gear calibration model, the random walk structure of the spatiotemporal field is first informed from a historical survey time series. Projections on stock abundance can be made based on the statistical properties of the random walk. Unlike a lag-1 autoregressive process, the mean of the spatiotemporal field does not approach zero, although the standard error increases (Thorson 2019; Sean Anderson, pers. comm.). With this behaviour, the model could estimate catchability of a new vessel based on the differences between the projected random walk and new survey data.

Model-based calibration was attempted for the 2019 trawl survey in 0A1CD in which the *Helga Maria* fishing vessel was contracted for the survey instead of the RV *Paamiut*. Although the Alfredo trawl gear was used on both vessels, it is believed that the gear interacted with the sea floor differently based on doorspread and trawl height (Nogueira and Treble 2020). Reduced contact of the trawl on the floor bottom would reduce catchability through fish outswimming or escaping the gear. It is also believed that the behavior of the Alfredo trawl on the FV *Helga Maria* relative to the RV *Paamiut* varied by depth.

The model is of the form:

$$\log \mu_i = X_{i,j}\beta_j + O_i + \omega_i + \varepsilon_i$$

where the design matrix $X_{i,j}$ included depth (as a quadratic term), a global intercept (corresponding to the catchability of the RV *Paamiut*), and additional vessel effects for the FV *Helga Maria* (2019 only). Three FV *Helga Maria* intercepts were included, corresponding to the catchability difference at three different depth strata (400–700 m, 701–1000 m, and 1000+ km). No year effect is included in the design matrix.

Year effects are implicit in the random walk structure of the spatiotemporal field $\varepsilon_{i,t}$ (an extra subscript indexes by time step t) where

$$\varepsilon_{i,t+1} = \varepsilon_{i,t} + \delta_{i,t}$$

$$\delta_{\cdot,t} \sim \mathrm{MVN}(0, \Sigma_\varepsilon)$$

All other variables are defined as in Appendix A.

Models were fitted separately for three different size classes based on the size-area distribution observed in the data. The likelihood of the model used the Tweedie distribution where the

variance is a power function of the mean, and provides more flexibility in the variance structure compared to the Poisson and negative binomial distributions.

The vessel effect was negative and significantly different at the $\alpha = 0.05$ level for the Small size category in the two deepest strata (701–1000, and 1000+ m) (Figure B.1, Table B.1). This is apparent in the 2019 catch rates containing values lower than the range of values seen in the RV *Paamiut* series (Figure B.2).

Once the model is fitted, the index is developed by predicting and summing the density over a spatial grid over 0A-South and 1CD (Figure B.3). By excluding the vessel effects in the prediction, the 2019 index is the value the *Paamiut* would have theoretically generated had the vessel performed the survey, although the confidence interval may be higher when there are no comparative gear studies.

Differences between the prediction grid of the spatiotemporal model and expansion factors of design-based estimates could explain the differences in scaling between the two indices (Figure B.3). Alternative configurations, for example, addition of a delta-model to incorporate presence/absence of the spatiotemporal model, could resolve scaling issues. The fitted model assumed that the encounter rate is one over the entire prediction grid which can contribute to higher abundance estimates. Finally, if index standardization were to use catch rates stratified by size, it would be desirable to fit a single model to account for correlations in catch rates and parameter estimates by size (rather than three separate models by size class). In this case, use of the spatiotemporal modeling facilitating multivariate regression available in the VAST software (Thorson 2019) is advised.

There were additional concerns regarding the timing of the 2019 survey, which occurred earlier in the season compared to prior years (Nogueira and Treble 2020), in which case seasonality would be confounded with the vessel effect. This confounding would be eliminated if the RV *Tarajoq* survey is held in the same season as the RV *Paamiut*. There were also concerns that too much time (five years) will have passed between the RV *Paamiut* and RV*Tarajoq* surveys to calibrate the index with sufficient precision. While it is possible to fit a model with one year of data from a new survey, it is not known whether the model will perform better with additional years of data.

Spatiotemporal modeling is a potential tool for model-based calibration in the absence of comparative towing. Simulation testing of the method is recommended and performance should be evaluated against the magnitude of various parameters: the spatial standard deviation (SD), the spatiotemporal SD, the time gap between surveys (relative to species longevity), and vessel effect. Case studies with comparative tow studies (e.g., Miller et al. 2010) can empirically validate this approach to estimate calibration factors.

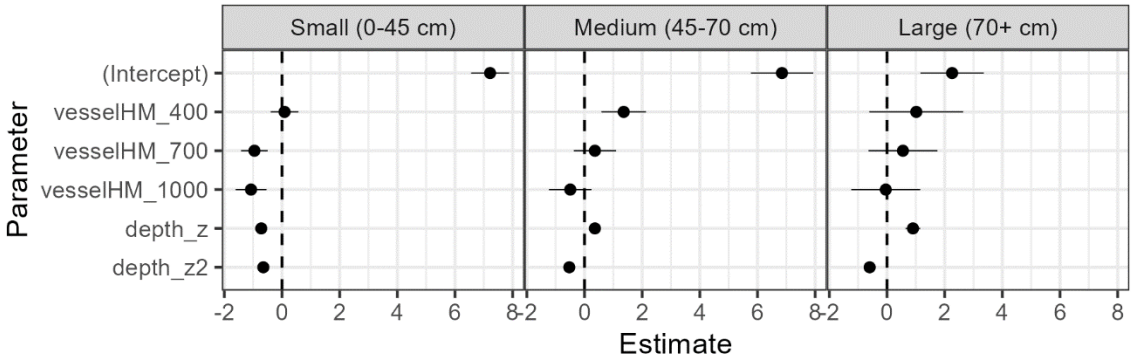

Figure B.1. Fixed effect coefficients from the sdmTMB model incorporating calibration of the 2019 FV Helga Maria *survey without comparative towing. Points indicate the maximum likelihood estimate and the line range indicates the 95 percent confidence interval using the standard error. Estimates are significantly different if the confidence interval does not include zero (dotted horizontal line). The intercept is interpreted as the catchability of the RV* Paamiut*. Three calibration factors are estimated for the FV* Helga Maria *(by depth strata of 400, 700, and 1000 m). Year effects are incorporated to the random walk of the spatiotemporal field and are not included here.*

Table B.1. Estimates of fixed effects (in log space) and nuisance parameters in the sdmTMB model incorporating calibration of the 2019 survey. Separate models were fit for survey catch rates for each size class (Small, Medium, Large). Asterisks indicate coefficients that were significant at the 95% level (null hypothesis of zero). `depth_z` is the tow depth (m) converted to a Z-score and `depth_z2` is the square of `depth_z`. Significance is not reported for nuisance parameters. `Tweedie p` is the exponent of the variance function of the mean.

Term	Estimate		
	Small	**Medium**	**Large**
(Intercept)	7.19*	6.82*	2.27*
vesselHM_400	0.14	1.31*	0.98
vesselHM_700	-0.92*	0.35	0.53
vesselHM_1000	-1.07*	-0.51	0.79
depth_z	-0.68*	0.37*	0.91*
depth_z2	-0.64*	-0.53*	-0.60*
Matérn range (km)	1386	2756	3473
Dispersion (ϕ)	6.95	6.82	5.15
Spatial SD	0.89	0.54	0.01
Spatiotemporal SD	0.29	0.47	0.67
Tweedie p	1.46	1.43	1.47

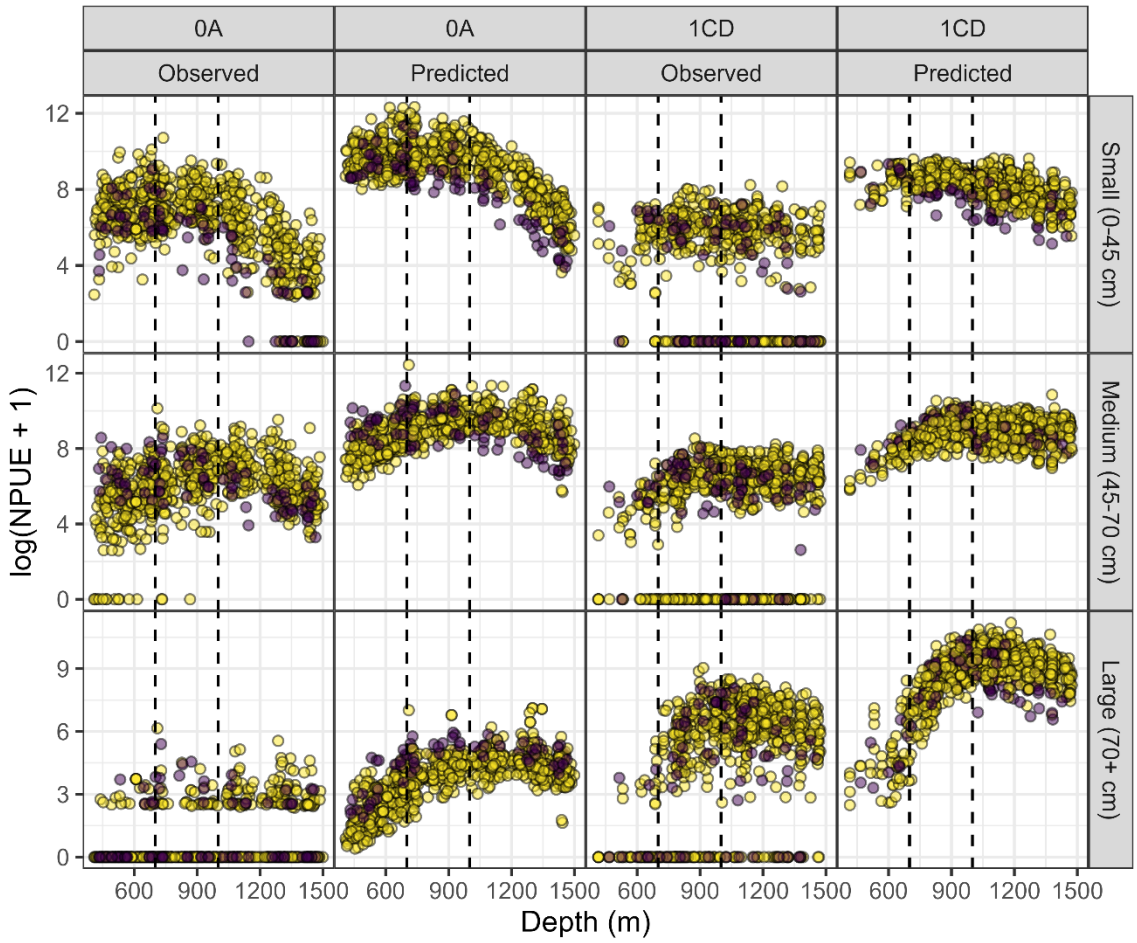

Figure B.2. Historical observed and model-predicted NPUE (numbers per sq. km swept area) by size class (Small, Medium, and Large) and depth in Division 0A-South and 1CD. Colors indicate the year and vessel of the set (purple for 2019 FV Helga Maria*, and yellow for RV* Paamiut *in all other years). The y-axis is transformed by taking the natural logarithm of NPUE + 1. The y-axis for the three size classes are not identical. Dotted vertical lines delineate the three depth strata for which the 2019 vessel calibration factor was estimated. Catchability differences are estimated in part by 2019 NPUE in some strata, e.g., for small fish in 0A, that is lower than the range observed in the RV* Paamiut *time series.*

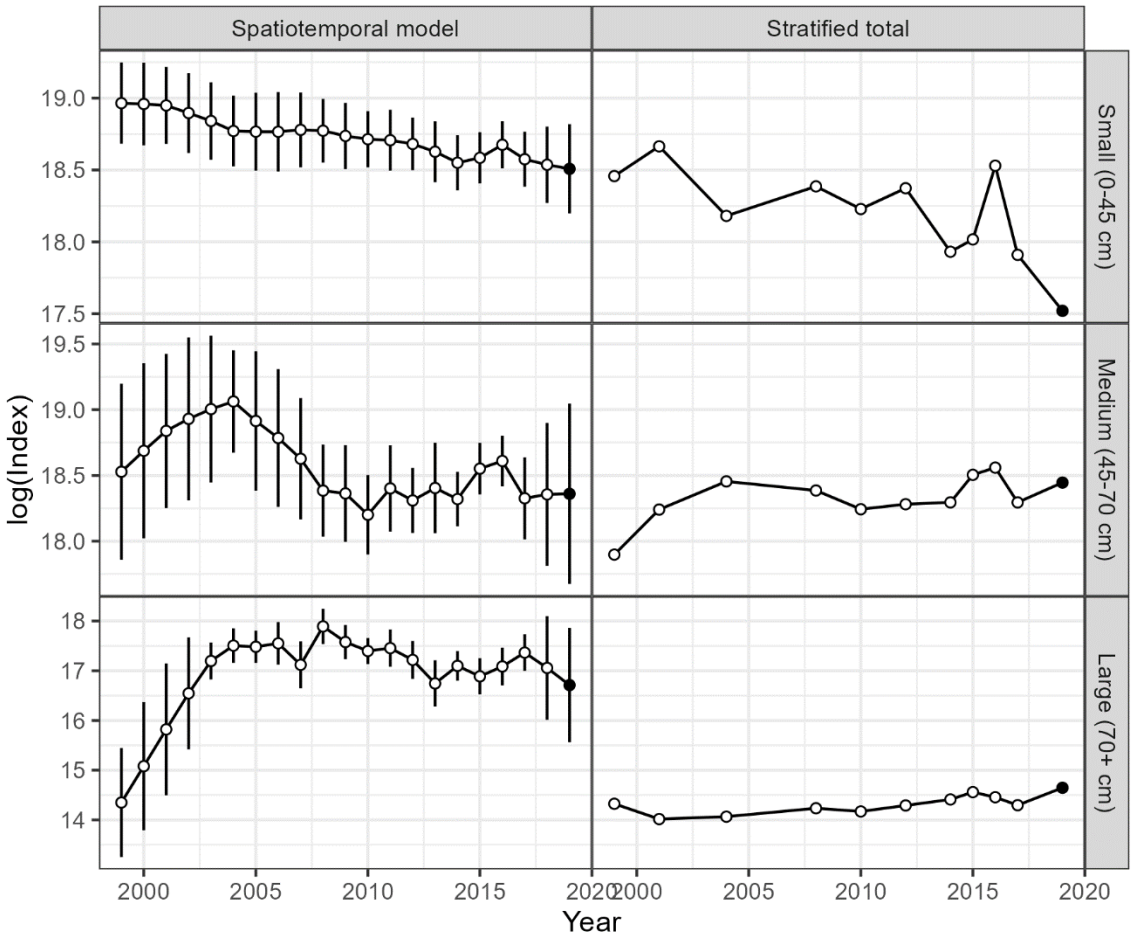

Figure B.3. Comparison of the natural logarithm of the index of abundance in 0A-South and 1CD derived from the spatiotemporal model and the stratified design-based estimate. For the spatiotemporal model, the 95% confidence interval is provided (not available for the design estimate). The design-based estimate can only be calculated when both 0A-South and 1CD are sampled in a given year, while the spatiotemporal model can impute in missing years and area strata if there is a random walk on the year effect. Index series include the 2019 survey (black point). In the spatiotemporal model, the 2019 estimate is the "RV Paamiut equivalent" after accounting for the estimated calibration factor for the FV Helga Maria. The confidence interval of the 2019 value is expected to be wider in the absence of comparative towing. The design-based estimate does not (and cannot) account for the catchability difference. Differences in scaling between the model-based and design-based index may be attributed to various factors, including (1) differences in the spatial grid used in the spatiotemporal model vs. the expansion factors in the stratified estimate, or (2) presence-absence assumption of the spatiotemporal model (assumed presence in all areas for all size classes). The spatiotemporal model estimate was generated by predicting the density in spatial cells in 0A-South and 1CD, see Figure 10 of the main text. No data from the 1CD survey prior to 2003 was used in the spatiotemporal model, which appreciably changes the index trend of the Large size class during these years.