

Formulating a spatio-temporal model to analyze longline survey data for the Atlantic Halibut fishery

Raphaël R. McDonald, Brad Hubley, Lingbo Li, Cornelia E. den Heyer,
and Joanna Mills Flemming

Science Branch, Maritimes Region
Fisheries and Oceans Canada
Bedford Institute of Oceanography
1 Challenger Drive
Dartmouth, Nova Scotia
B2Y 4A2

2023

**Canadian Technical Report of
Fisheries and Aquatic Sciences #####**



Fisheries and Oceans
Canada

Pêches et Océans
Canada

Canada

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. Il 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 nnn

2023

FORMULATING A SPATIO-TEMPORAL MODEL TO ANALYZE LONGLINE SURVEY DATA FOR
THE ATLANTIC HALIBUT FISHERY

by

Raphaël R. McDonald¹, Brad Hubley², Lingbo Li², Cornelia E. den Heyer², and Joanna Mills
Flemming¹

¹Department of Mathematics and Statistics
Dalhousie University, 6299 South Street
Halifax, Nova Scotia, B3H 4R2, Canada

²Science Branch, Maritimes Region
Fisheries and Oceans Canada
Bedford Institute of Oceanography
1 Challenger Drive, PO Box 1006
Dartmouth, NS, B2Y 4A2

© His Majesty the King in Right of Canada, as represented by the Minister of the
Department of Fisheries and Oceans, 2023
Cat. No. Fs97-6/nnnE-PDF ISBN ISSN 1488-5379

Correct citation for this publication:

McDonald, R.R., Hubley, B., Li, L., den Heyer, C.E., and Mills Flemming, J. 2023. Formulating a spatio-temporal model to analyze longline survey data for the Atlantic Halibut fishery. Can. Tech. Rep. Fish. Aquat. Sci. nnn: vi + 36 p.

CONTENTS

ABSTRACT	v
RÉSUMÉ	vi
1 Introduction	1
2 Methods	2
2.1 Study Area	2
2.2 Survey designs	3
2.2.1 Fixed Stations Design	3
2.2.2 Stratified Random Design	4
2.3 Model Formulation	4
2.3.1 MEM	4
2.3.2 MEMSpa	5
2.3.3 Spatio-Temporal MEM	6
2.4 Data	8
2.5 Persistence of spatial patterns	8
3 Results	9
3.1 Model Selection	9
3.2 Model Validation	10
3.3 Fits to Alternative Data Subsets	10
3.4 Persistence	12
4 Discussion	12
5 Acknowledgements	14
6 References	14

APPENDICES	16
A	24
B	26
C	34
7 References	36

ABSTRACT

McDonald, R.R., Hubley, B., Li, L., den Heyer, C.E., and Mills Flemming, J. 2023. Formulating a spatio-temporal model to analyze longline survey data for the Atlantic Halibut fishery. Can. Tech. Rep. Fish. Aquat. Sci. nnn: vi + 36 p.

The Industry-DFO Atlantic halibut (*Hippoglossus hippoglossus*) longline survey has been running since 1998 (using a fixed stations design) with the aim of providing a reliable index of halibut abundance. **A 2016 review of modelling approaches recommended the use of a multinomial exponential model (MEM) to account for hook competition and the abundance of non-target species and further to survey redesign. A new stratified random survey was initiated in 2017.** Since then, this new stratified survey was run in parallel with the fixed station survey, but with a reduced number of fixed stations. The new survey has more standardized fishing protocols and additional hook condition data are collected to inform the MEM. **The aim of this report is to present a new spatio-temporal version of MEM that is able to successfully model data from both the fixed stations and the stratified random sampling design within a unified framework. It also incorporates additional fixed and random effects, where the random effects account for the among-group variability associated with fishing vessels.** The analysis demonstrates the potential bias in abundance estimates associated with modelling changes in halibut and non-target distributions and abundance over time that would occur using only the fixed stations design. It further shows that the switch to a stratified random sampling design was timely and helped better quantify the increase in halibut catch rates observed since 2017.

RÉSUMÉ

McDonald, R.R., Hubley, B., Li, L., den Heyer, C.E., and Mills Flemming, J. 2023. Formulating a spatio-temporal model to analyze longline survey data for the Atlantic Halibut fishery. Can. Tech. Rep. Fish. Aquat. Sci. nnn: vi + 36 p.

1 Introduction

Survey data for the Atlantic halibut (*Hippoglossus hippoglossus*) population on the Scotian Shelf were originally obtained as part of the Fisheries and Ocean's (DFO) summer Research Vessel (RV) bottom trawl survey (den Heyer et al. 2015). This survey provides indices of abundance, indices of biomass, and biological samples for groundfish and other components of the ecosystem. Some commercially valuable fish and invertebrate stocks are not well sampled by trawl gear, such as halibut (Zwanenburg et al. 2003), and alternate methods to monitor abundance are used. In the case of halibut, the Industry-DFO Halibut longline survey was initiated in 1998 using longline gear to provide indices of abundance. A fixed stations design was adopted and fishing was completed by commercial fishermen, with at-sea observers recording data on the catch by weight of target species (Atlantic halibut) and non-target species (anything else) (Zwanenburg and Wilson 2000; Trzcinski et al. 2009; den Heyer et al. 2015). The aim of this new survey was to obtain information on size composition, diet, movement and bycatch species for stock assessment purposes (Zwanenburg and Wilson 2000). Initially, the stratified mean catch per hook was used as an index of abundance. In 2009, a generalized linear model was implemented to obtain standardized survey catch rates of halibut weight (Trzcinski et al. 2009; Smith 2016).

An extensive review of methods to estimate catch per unit effort for longline gear (Smith 2016) concluded that catch per unit effort standardization methods ignored the presence and abundance of non-target species and both intra- and inter-species competition for hooks. A multivariate Multinomial Exponential Model (MEM) originally developed by Rothschild (1967) and later reformulated by Etienne et al. (2013) was recommended. The switch to this model required hook condition data, such as the number of baited and unbaited hooks, in addition to total numbers of target and non-target species caught (Smith 2016). In 2017, a new stratified random survey was established with more standardized protocols and strata that encompassed the entire management unit (DFO 2021). To obtain the data for the MEM model, this new survey also required the collection of hook condition data. To calibrate the new stratified random survey, a subset of fixed station sets have also been completed to provide the opportunity to calibrate the surveys.

While the MEM showed great promise and could easily be modified to incorporate covariates (Smith 2016), its original formulation could not account for any spatial patterns in catch rates (Luo et al. 2022) which are likely to exist due to the scale and environmental complexity of the Atlantic halibut management areas (Shackell et al. 2021). Recent work utilizing geostatistics was able to harness this previously ignored information to obtain improved estimates of both station-specific and overall catch rates (Luo et al. 2022). However, this work focused exclusively on the data from the stratified random sampling design (2017 onwards), leaving 19 years of data unused. Given the importance of this survey for halibut management, the Atlantic Halibut Council sponsored the extension of this strictly spatial approach to a fully spatio-temporal model that could include data from both surveys to provide an index from 1998 to the present. The Term of Reference for this work (including the agreed-upon deliverables) are provided in Appendix 1. This report aims to present a fully spatio-temporal MEM and a comparative analysis when utilizing data from the Scotian Shelf and the Southern Grand Banks (NAFO divisions 3NOP4VWX) on the eastern coast of Canada. It further aims to assess the ability of only the fixed stations data to account for temporal changes in the abundance of Atlantic halibut.

2 Methods

2.1 Study Area

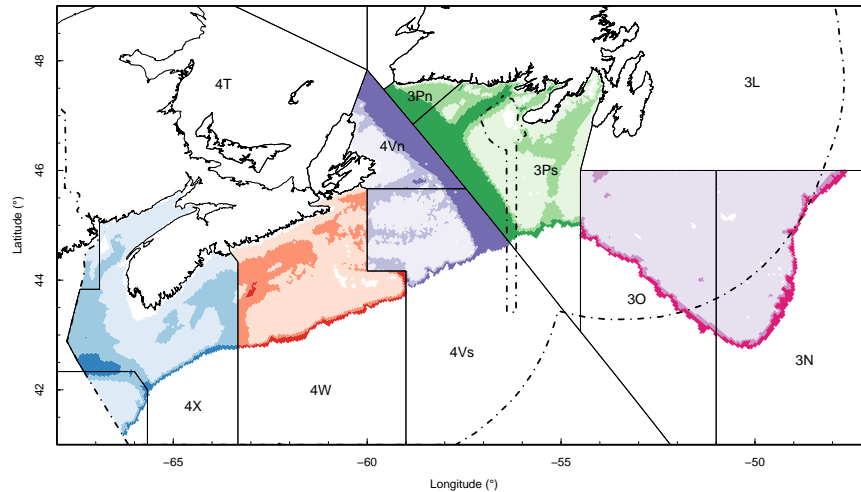


Figure 1. NAFO divisions on the Scotian Shelf and the Southern Grand Banks. The 3NOPs4VWX5Zc Atlantic halibut stock survey area is denoted by the coloured area, with five area strata: 4X5YX (blues), 4W (oranges), 4V (purples), 3P (greens), and 3NO (reds), and three depth strata: 30 to 130 m (light colour), 131 to 250 m (medium colour), and 251 to 750 m (dark colour). NAFO subdivisions are labelled, and separated by solid lines. The exclusive economic zones of Canada and France are shown with dashed lines. NAFO subdivision 3Pn is not part of the stock area, but has been included in the survey since 2017.

There are 2 Atlantic Halibut management units in Canada. The focus of this work is the Scotian Shelf and Southern Grand Banks management unit which encompasses the North Atlantic Fisheries Organization (NAFO) divisions 3NOPs4VWX5Zc. The other NAFO division 4RST encompasses the Gulf of St. Lawrence and will not be included in this analysis. The definition of these management units was mostly based on tagging studies that showed Atlantic halibut moving throughout most of the Canadian North Atlantic (DFO 2021), and subsequently has been supported by recent genetic analysis (Kess et al. 2021). Halibut is caught throughout the management unit and on the tail of the Grand Banks outside of Canada's exclusive economic zone (EEZ), mostly along the continental shelf (DFO 2021), see Figure 1.

Due to the nature of spatial modelling and our desire to scale station-specific indices up to an overall index representing the whole area, the area modelled is bounded by the continental shelf (750 m contour) as shown in Figure 1. Since no survey ever extends outside of these edges and we do not wish to extrapolate outside the bounds of the observations, restricting our model to this area was determined to be appropriate.

2.2 Survey designs

As mentioned previously, there are 2 surveys with different sampling designs contributing to this analysis: a survey that follows a fixed stations design (1998 to 2021), and a survey following a stratified random sampling design (2017 to 2021).

2.2.1 Fixed Stations Design

Initially, the formulation of this survey followed a stratified design with 222 fixed stations, where 30 were reserved for the 3NOPs subdivision (den Heyer et al. 2015). The strata were defined based on observed halibut landings by trips between 1993 and 1997 using 3 categories: high catches (>250 kg), medium catches (50-249 kg) and low catches (<49 kg) with the number of planned stations proportionally allocated following a ratio of 5:7:10 for the low, medium, and high strata respectively (Zwanenburg and Wilson 2000; Zwanenburg et al. 2003; den Heyer et al. 2015; Smith 2016). Station coverage has been inconsistent over time as not all stations have been covered every year and new stations were added in the mid-2000s (den Heyer et al. 2015). The number of fixed stations decreased to around ~100 per year in 2017 due to the implementation of a new stratified random survey. See Table 1 for the number of stations sampled every year.

The survey fishing protocol was to set 1000 hooks (Mustad circle hooks #14 or greater) per station and soak for 10 hours between 4 am and noon. Variations in both number and size of hooks (size #16 hooks becoming more common later in the time series) did occur (den Heyer et al. 2015). The start or end of the longline was supposed to be within 3 nautical miles of the station. Vessel and captain participation varied over time due to many different factors, with 66 different vessels sampling the fixed stations since 2000. The number of halibut (including sub-legal size fish) and non-target species were counted in each set.

Table 1. Number of successfully sampled stations by sampling design between 1998 and 2021.

Year	Fixed Stations	Stratified Stations	Total Stations
1998	174	N/A	174
1999	166	N/A	166
2000	216	N/A	216
2001	190	N/A	190
2002	199	N/A	199
2003	188	N/A	188
2004	215	N/A	215
2005	164	N/A	164
2006	163	N/A	163
2007	241	N/A	241
2008	281	N/A	281
2009	205	N/A	205
2010	215	N/A	215
2011	217	N/A	217
2012	217	N/A	217

Year	Fixed Stations	Stratified Stations	Total Stations
2013	233	N/A	233
2014	232	N/A	232
2015	232	N/A	232
2016	227	N/A	227
2017	98	141	239
2018	100	150	250
2019	98	123	221
2020	99	148	247
2021	98	131	229

2.2.2 Stratified Random Design

The survey follows a stratified random sampling design, where the strata are the 5 NAFO subdivisions (4X5YZ, 4W, 4V, 3P, 3NO) each with 3 depth strata (30-130 m, 131-250 m, 251-750 m) and includes 3Pn and some areas outside of Canada's EEZ, which were not part of the management unit (Cox et al. 2018; Luo et al. 2022). The depth bounds (30-750 m) based on exploratory analyses of catch rates by depth from the fixed stations (Cox et al. 2018), contained most of the survey sets as well as most of the Atlantic halibut habitat. Approximately 150 survey stations per year were randomly assigned to strata with the number in a given stratum proportional to its size (Cox et al. 2018; Luo et al. 2022). The fishing protocol was similar to the fixed stations protocol but notably more strictly defined, with each set containing 1000 baited hooks (size #15 hooks) set for between 6 and 12 hours (Luo et al. 2022). Forty different vessels have participated in the stratified survey sampling design, with 19 of those also sampling some of the fixed stations at some point (meaning 21 vessels have only participated in the stratified random survey).

Unlike the fixed stations, the observers collecting the stratified random survey data also recorded hook condition data on a subset of hooks for each set. Based on a pilot study that aimed to calculate the quantity of hooks required to be broadly representative of the whole set (Doherty et al. In press), 10 samples of 30 hooks across the line were chosen for a total of 300 per 1000 hooks (Luo et al. 2022). Instead of simply counting the number of halibut (including sub-legal size fish) and non-target species caught, the condition of each hook (baited, unbaited, broken, missing) is also recorded for this subsample. See Table 1 for the number of stations sampled every year.

2.3 Model Formulation

2.3.1 MEM

The original aim of the MEM was to account for hook competition in longline fishing as proposed by Rothschild (1967) and reformulated by Etienne et al. (2013). There are two different formulations of the MEM, which will respectively be called the Full MEM and the Reduced MEM.

The Reduced MEM allows for three possible outcomes for every hook following the summation $N_i = N_{B,i} + N_{T,i} + N_{NT,i}$, wherein the total number of hooks N_i on a longline set i is the sum of the number of hooks with non-target species $N_{NT,i}$, the number of hooks with target species $N_{T,i}$, and the number of hooks without any animals that are assumed to still be baited $N_{B,i}$. Assuming that the time to catch a target or non-target species follows independent exponential distributions with rates λ_T and λ_{NT} , then the vector $(N_{B,i}, N_{T,i}, N_{NT,i})$ follows a multinomial distribution (Etienne et al. 2013; Luo 2020; Luo et al. 2022):

$$(N_{B,i}, N_{T,i}, N_{NT,i}) \sim \mathcal{M}(N_i, \alpha_i) \text{ where} \quad (1)$$

$$\alpha_i = (e^{-\lambda S_i}, (1 - e^{-\lambda S_i}) \frac{\lambda_T}{\lambda}, (1 - e^{-\lambda S_i}) \frac{\lambda_{NT}}{\lambda}), \quad (2)$$

S_i is the soak time of longline set i , and $\lambda = \lambda_T + \lambda_{NT}$.

This formulation does not account for the condition of hooks that return without catching animals, as these are not guaranteed to still be baited. Etienne et al. (2013) therefore reformulated this MEM into the Full MEM, wherein hooks can also come back unbaited (which includes broken or missing hooks). These empty hooks are assumed to come from interactions with animals, but due to identifiability concerns another assumption has to be made as to whether they are caused by target or non-target species (Etienne et al. 2013). Since the gear is more likely to be oriented towards catching target species and that non-target species are likely more abundant (encompassing many more species than just the main target), it seemed more reasonable to assume that empty or broken hooks were caused by non-target species (Etienne et al. 2013; Luo et al. 2022). Furthermore, while this is likely to lead to a negative bias in the estimated catch rates, it is the most conservative choice as it guarantees there will be no positive bias caused by the assumption of escapes only caused by target species or target and non-target species having the same probability of escape (see results when this assumption is made in Appendix B). This extra outcome is therefore added to the multinomial model, which now follows the following distribution:

$$(N_{B,i}, N_{T,i}, N_{NT,i}, N_{E,i}) \sim \mathcal{M}(N_i, \alpha_i) \text{ where} \quad (3)$$

$$\alpha_i = (e^{-\lambda S_i}, (1 - e^{-\lambda S_i}) \frac{\lambda_T}{\lambda}, (1 - e^{-\lambda S_i}) \frac{\lambda_{NT}}{\lambda} (1 - p_{NT}), (1 - e^{-\lambda S_i}) \frac{\lambda_{NT} p_{NT}}{\lambda}), \quad (4)$$

$N_{E,i}$ is the number of empty hooks, and p_{NT} is the probability of escape of non-target animals. In this formulation, p_{NT} shows up in the calculation of $N_{NT,i}$, as it impacts the catch rates of non-target species. As this component is present in the Reduced MEM and those non-target species would be similarly able to escape as in the Full MEM, we modified the Reduced MEM so that the multinomial component $N_{NT,i}$ is obtained as $(1 - e^{-\lambda S_i}) \frac{\lambda_{NT}}{\lambda} (1 - p_{NT})$. While it is unlikely that this version of the Reduced MEM can reliably estimate p_{NT} , it should be able to borrow the information from the Full MEM when both datasets are analyzed in the same framework.

2.3.2 MEMSpa

While the MEM model achieved its goal of incorporating hook competition to improve estimates of catch rates, it did not account for spatial patterns in the survey data. Previous work (Luo 2020; Luo et al. 2022) incorporated geostatistical approaches to modify both versions of the MEM through the use of Gaussian Random Fields (GRF).

The observation level of this new MEMSpa remained almost the same as described in the previous section, but the rates λ_T and λ_{NT} were modified to incorporate the location of a given longline set i :

$$\lambda_{T,i} = \exp(\beta_T + \omega_{T,i}) \quad (5)$$

$$\lambda_{NT,i} = \exp(\beta_{NT} + \omega_{NT,i}) \quad (6)$$

where $\lambda_{T,i}$ and $\lambda_{NT,i}$ are the exponential rates for target and non-target species at the location of longline set i , β_T and β_{NT} are intercept parameters, and $\omega_{T,i}$ and $\omega_{NT,i}$ are the values of the underlying GRF.

These modifications allow the model to incorporate spatial patterns present in the data to obtain station-specific catch rates. An additional requirement is to be able to obtain a single overall estimated rate for the entire modelled area to be treated as an index of relative abundance. Due to the computational load of utilizing kriging to obtain this index, a Dirichlet method is preferred wherein the modelled area is divided into disjoint tiles based on the survey station locations, and each station is associated with a specific region and assumed to be representative of it (Luo et al. 2022). One can then obtain a spatially-weighted survey index for the whole area as follows:

$$OverallIndex = \frac{\sum_{i=1}^I A_i \hat{\lambda}_i}{\sum_{i=1}^I A_i} \quad (7)$$

where A_i is the area of the Dirichlet tile associated with station i , $\hat{\lambda}_i$ is the corresponding estimated catch rate at this station, and I is the total number of stations. An alternative general index could also be considered in the yearly intercepts, as these would represent the expected catch rate after accounting for the effect of space and vessels.

As the data from the stratified random survey only has a subset of its hooks where the hook condition was recorded, a product likelihood approach was taken so as to incorporate all the available data inside a unified framework. This consists of separately fitting the Reduced MEMSpa to the data without hook condition and the Full MEMSpa to the data with hook condition, and multiplying their likelihoods together (see Luo et al. (2022) for more details on MEMSpa).

2.3.3 Spatio-Temporal MEM

The improvements brought forward by the inclusion of spatial patterns in MEMSpa were very clear when fitted to the stratified dataset using the product likelihood method to combine both the 300 and 700 hook subsets, but there still remained years of available information from the fixed stations that had not been analyzed. Furthermore, MEMSpa could only be fit to a single year's data at a time and hence did not account for any temporal patterns in halibut or non-target species distributions and abundance. Incorporating the spatio-temporal variability in catches in the MEMSpa would result in a fully spatio-temporal MEM.

As we aim to retain the inclusion of spatial patterns through the residual structure as done in MEMSpa, we chose to incorporate the temporal aspects in the mean structure. There are many different approaches for incorporating temporal patterns in the mean and these usually involve random effects, with common approaches including random intercepts (Venables and Dichmont

2004; Kai et al. 2017; Pedersen et al. 2018), random slopes (Venables and Dichmont 2004; Swain et al. 2009), random walks (Swain et al. 2009; Li et al. 2019), or autoregressive models (Schnute and Richards 1995; Kai et al. 2017). These last two could also be seen as more structured versions of other approaches wherein the random walk or autoregressive structure would be on the random intercept. Furthermore, we also decided to account for variability among different survey vessels by including random vessel effects. While it is likely that there will be some confounding between vessel effects and spatial and temporal effects because some vessels will be restricted to specific years or areas, including them should still help separate the impact of most vessels from the effects of interest. The novel formulation takes the following form:

$$\lambda_{T,i,y} = \exp(\eta_{T,y} + \nu_{T,j} + \omega_{T,i,y}) \quad (8)$$

$$\lambda_{T,i,y} = \exp(\beta_T + \eta_{T,y} + \nu_{T,j} + \omega_{T,i,y}) \quad (9)$$

where $\lambda_{T,i,y}$ is the exponential rate for the target species at station i in year y , $\nu_{T,j}$ is the random effect of vessel j ($\nu_{T,j} \sim N(0, \sigma_\nu^2)$), $\eta_{T,y}$ is the random intercept for target species in year y or, in Equation 9, the random slope in year y with global intercept β_T . $\omega_{T,i,y}$ are the GRF values for target species at station i in year y . In Equation 8, $\eta_{T,y} \sim N(\mu_{T,int}, \sigma_{T,int}^2)$, while for Equation 9 $\mu_{T,int} = 0$. For the random walk and autoregressive models, they follow the same formulation as Equation 8 with the following added structure on $\eta_{T,y}$:

$$\eta_{T,y} = \eta_{T,y-1} + \epsilon_y, \quad \epsilon_y \sim N(0, \sigma_\eta^2) \quad (10)$$

$$\eta_{T,y} = c + \phi\eta_{T,y-1} + \epsilon_y, \quad \epsilon_y \sim N(0, \sigma_\eta^2) \quad (11)$$

where ϵ_y is an error term ($\epsilon_y \sim N(0, \sigma_\eta^2)$). The autoregressive model in Equation 11 is a first order autoregressive (AR(1)) process with constant mean c and autoregressive parameter ϕ bounded between -1 and 1 to ensure stationarity. Non-target rates follow the same structure as above with non-target specific parameters.

These models are fit to the data from both surveys. As the fixed stations data do not contain any information on hook condition, only the Reduced MEM can be fit to it. Model selection will be done by comparing root mean squared errors (RMSE), and by calculating AIC and BIC values for the 4 different model fits to every data subsets examined. These subset include the following: the stratified data (2017 to 2020), the fixed stations (1998 to 2020), both datasets in overlapping years (2017 to 2020) and both datasets for all years (1998 to 2020). There are 2 types of errors to consider in this model, one being the GRF at the hierarchical level of the catch rates λ and the other being the difference between the observations and the expected observations based on the probabilities obtained from the multinomial distribution. Either way, the RMSE is calculated as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N e_i^2}{N}} \quad (12)$$

where e_i is the error, either the GRF value $\omega_{i,y}$ for rate i in year y or the total difference between the observed and expected value of target, non-target and hooks for a given longline set i .

AIC (Akaike 1974) and BIC (Schwarz 1978) are both information criterion that are used to compare different models fit to the same data and try to balance between better model fit (through higher likelihood) and number of parameters. They are respectively calculated as:

$$AIC = 2K - 2\log(L) \quad (13)$$

$$BIC = K\log(n) - 2\log(L) \quad (14)$$

where K is the number of parameters, $\log(\cdot)$ is the natural log, L is the maximized likelihood, and n is the number of data points.

Using the best model identified for the fit with both datasets from 2000 to 2021, a 10-fold cross-validation is performed to test the predictive ability of this model. This means that both datasets are split into 2 portions containing 90% (training set) and 10% (test set). The model is then fit to the training set to estimate parameters and predict random effects, after which the model output is used to predict the data from the test set. Ordinary kriging is used to obtain the GRF values at the locations of the test sets. The differences between the predictions of the test sets and the actual observations can then be observed using out-of-sample RMSE calculated as before.

2.4 Data

Models were fit to different data subset: the stratified data (2017 to 2020), the fixed stations (1998 to 2020), both datasets in overlapping years (2017 to 2020) and both datasets for all years (1998 to 2020). For comparative purposes, the nominal catch rate in each year is calculated by dividing the observed number of halibut caught at each station by the product of the number of hooks and the soak time of each station. This was done separately for both datasets combined and for the fixed station dataset. Code used to analyze the data is available at <https://github.com/RaphMcDo/TechRepMEM/tree/master/Analysis%20Scripts>.

2.5 Persistence of spatial patterns

The ability of the fixed stations data to accurately capture changes in the relative abundance of Atlantic halibut and non-target species abundance was explored. Generally, obtaining an unbiased estimate of population abundance when only using fixed stations is very difficult due to the absence of reliable design-based estimators (Li et al. 2015; Lee and Rock 2018). Furthermore, an underlying assumption of a fixed station survey design is that the spatial distribution of the population of interest is consistent over time (Li et al. 2015; Lee and Rock 2018), as changes in this spatial distribution would lower the ability of this design to track changes over time and reduce the efficiency of the strata used for the original choice of stations. In our case, this means that the distribution of Atlantic halibut and non-target species would have to be consistent with their spatial distribution between 1995 and 1997. The fixed stations design was able to detect changes in abundance (den Heyer et al. 2015; Trzcinski and Bowen 2016; Li et al. 2022), but there was concern that catch rates might not be proportional to abundance (Smith 2016; Cox et al. 2018). As this stock recovered, the distribution of the commercial fishery changed and expanded (den Heyer et al. 2015; Li et al. 2022), highlighting the need for survey coverage throughout the management unit.

A relatively straightforward approach to test the fixed stations for their ability to appropriately track population abundance change over time is test the persistence of these distributions over time (Lee and Rock 2018). Given that the counts transformed to catch rates violated the basic assumptions of the traditional ANOVA approach chosen by Lee and Rock (2018) (as our data was not normally distributed), we evaluated persistence by conducting a series of pairwise

comparisons of halibut and non-target species catch rates between years. The statistic was calculated as follows (Warren 1994; Li et al. 2015):

$$\bar{\omega} = \frac{s_y^2/4}{s_s^2 - s_y^2/4} \quad (15)$$

where $\bar{\omega}$ is the measurement of the degree of persistence where a smaller value indicates a greater degree of persistence, s_y^2 is the difference in catch rates of the same site between compared years, and s_s^2 is the difference in catch rates between different sites in the same year. These last 2 variables are calculated as:

$$s_s^2 = \frac{\sum_{y=1}^2 \sum_{i=1}^{n_i} (x_{iy} - \bar{x}_y)^2}{m_1 + m_2 - 2} \quad (16)$$

$$s_y^2 = \sum_{i=1}^m (d_i - \bar{d})^2 / (m - 1) \quad (17)$$

where x_{iy} is the observed catch rate in site i and year y , \bar{x}_y is the mean observed catch rate of year y , m_1 is the number of fixed stations in the first year included in the pairwise comparison while m_2 is the same for the second year, d_i is the difference in catch rate between two years in site i and \bar{d} is the mean catch rate difference. An important note here is that, as the coverage of stations changed over time, there are often more stations included in the s_s^2 than s_y^2 as all stations in both years are included in s_s^2 , but only the stations fished in both years can be included in s_y^2 . Since no p-values are used in this approach, no explicit multiple comparison corrections (e.g., Bonferonni correction) can be performed, but the large number of pairwise comparisons should temper any conclusions drawn from this analysis. Furthermore, this analysis is meant as a quick exploration to help identify potential reasons why the Spatio-Temporal MEM might identify different patterns when fit to both datasets or just the fixed station data.

3 Results

3.1 Model Selection

Due to issues fitting the 1999 data to all model formulations, the analysis focuses on the data starting in 2000 and does not incorporate the data from 1998 and 1999.

All 4 models converged for every data subset tested here. The RMSE, AIC and BIC values from this model validation process can be seen in Table 2. These values appear to indicate that there is little difference between the model fits when it comes to RMSE. According to the AIC and BIC values, the best model would be the random walk as it has the lowest values for both metrics. This was the case for every single fit with the exception of the fit using both datasets from 2017 to 2021, where the random intercept and random slope models were equal in their AIC and BIC values (see Appendix C). However, closer analysis of both fits showed that the Hessian matrices were not positive-definite, which made it unable to estimate the standard error of some variance parameters and potentially indicating issues with convergence. Rejecting those models, the next best model was the random walk model as for the other fits. The model used for all analysis moving forward is therefore the random walk model.

Table 2. Outputs for model selection approaches when fitting random walk model to both datasets from 2000 to 2021, including root mean squared errors (RMSE), Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC).

Model	Observation RMSE	Target Field RMSE	Non-Target Field RMSE	AIC	BIC
Random Intercept	0.6711552	1.310907	1.705889	230,747.7	230,819.0
Random Walk	0.6711413	1.301179	1.712035	230,686.6	230,757.8
Random Slope	0.6711552	1.310907	1.705885	230,747.7	230,819.0
AR(1) Process	0.6711544	1.311352	1.706027	230,750.3	230,834.5

3.2 Model Validation

Table 3. RMSE (in predicted numbers) and RMSE standard error obtained from 10-fold cross-validation output using the random walk model to both datasets from 2000 to 2021

RMSE Type	RMSE	SE
Overall RMSE	43.185422	1.5553442
Target RMSE	5.564864	0.4198784
Non-Target RMSE	56.908617	3.0792351

The RMSE calculated from the 10-fold cross-validation can be seen in Table 3. In this case, the Overall RMSE aggregates the RMSE for target species, non-target species, and empty hooks (including baited and unbaited hooks in the stratified dataset) and therefore represent the total number of hooks with the wrong outcome (e.g., 1 hook predicted to have a target species when it was actually empty) when predicting on the test dataset. The RMSE for target species, the main variable of interest for this model, is fairly small as it represents ~0.5% of the average number of hooks (1000). This is not surprising as the target species usually represents the smallest component of a given set, as there are usually more non-target species and empty hooks than halibut. An important caveat here is that 2 of the fits resulted in false convergence, but their RMSE and parameter estimates were not substantially different from the other folds. As the datasets have a substantial amount of zeroes and present complex spatio-temporal patterns, it is not entirely surprising that it would be difficult to work with when some data is removed. There were no obvious differences in the training data used for these specific folds with the only potential explanation that the peak observed number of halibut was higher in the test datasets for these folds, which might indicate that these strong tows contain a lot of information from the perspective of the model. Overall, there were no substantial differences in the RMSE calculated for each fold (see Appendix C).

3.3 Fits to Alternative Data Subsets

As the best model was considered to be the random walk approach for all data combinations, that model was the one fit to the varying data subsets. As for model selection, data from 1998 and 1999 were not included in the analysis.

The model successfully converged when fit to all data subsets, but the fits to the subsets with data from 2017 to 2021 showed signs of overparameterization as all of them have at least 1 variance parameter related to the random intercepts that is estimated arbitrarily close to 0 (see Table 4). Most other parameters are estimated very similarly for all fits with the exception of p_{nt} (the probability of escape for non-target animals) and μ_{nt} (the mean of the random intercept distribution for 2000). p_{nt} is estimated at a very low value when only the fixed stations are included with a correspondingly lower estimate of μ_{nt} , while estimates based on fits to the stratified data are substantially higher for both parameters.

Table 4. Estimated values of parameter for each different model fit with standard error in parentheses.

Parameter	Both Datasets (Full)	Fixed Stations (Full)	Both Datasets (2017+)	Fixed Station (2017+)	Stratified Data (2017+)
p_{nt}	0.788 (0.001)	0.129 (0.009)	0.881 (0.0005)	0.246 (0.015)	0.903 (0.0005)
ϕ_t	0.050 (0.002)	0.053 (0.002)	0.054 (0.005)	0.060 (0.007)	0.058 (0.007)
ϕ_{nt}	0.048 (0.001)	0.060 (0.002)	0.015 (0.002)	0.046 (0.005)	0.051 (0.005)
σ_t	1.619 (0.032)	1.468 (0.032)	1.956 (0.085)	1.408 (0.080)	1.985 (0.101)
σ_{nt}	1.915 (0.030)	1.599 (0.029)	1.493 (0.038)	1.582 (0.074)	0.885 (0.029)
$\sigma_{vess,t}$	0.760 (0.099)	0.601 (0.083)	0.735 (0.228)	0.485 (0.174)	0.505 (0.214)
$\sigma_{vess,nt}$	0.882 (0.091)	0.744 (0.088)	0.746 (0.108)	0.748 (0.187)	0.624 (0.092)
$\sigma_{int,t}$	0.114 (0.038)	0.185 (0.048)	0.0003 (0.043)	0.00010 (0.083)	0.00005 (0.030)
$\sigma_{int,nt}$	0.293 (0.074)	0.078 (0.058)	0.158 (0.119)	0.233 (0.220)	0.0002 (0.038)
μ_t	-12.590 (0.221)	-12.866 (0.264)	-12.012 (0.181)	-11.133 (0.174)	-12.262 (0.164)
μ_{nt}	-7.531 (0.384)	-9.432 (0.202)	-6.319 (0.225)	-9.559 (0.381)	-6.015 (0.117)

The distribution of vessel effects are very similar between model fits, with more variability and a wider distribution when more data is incorporated (i.e., fits with just the fixed stations or just the stratified data from 2017 onwards tend to have lower variance and tighter distribution, see Figure 2). This can be caused by many factors including the stricter protocol for the stratified random sets that could result in less variability between vessels or differences between individual vessels and the areas and years they fished. The overparameterization of fits to data subsets that only contain data from 2017 onwards is very visible in Figure 3 where most intercepts do not vary annually due to the model not being able to estimate the variance around them (the uncertainty around them represents the uncertainty around their estimated mean parameters, which are well estimated). For the fits that have data starting in 2000, there are clear changes over time in the random intercepts. Non-target species intercepts are always lower when only the fixed stations are considered, as expected with the difference in the estimated value of p_{nt} . Notably, the stratified random survey, with the additional hook condition data, suggests an increase in overall average expected non-target catch rates starting in 2016 which does not show up when the model is fit only to fixed station data.

The pattern for the target species (halibut) are a lot more similar between model fits, with both capturing an overall increase in expected catch rates over the time series. However, the fit with just the fixed stations has a slightly larger increase. For the indices themselves (λ_t and λ_{nt}), there are large differences when looking at the fits that only include data from 2017 onwards (Figure 4). The fit with just the fixed stations has a completely different pattern than the other 2 fits. Those other 2 fits have similar trends, but the estimates when both datasets are included are more than twice as large as when just the stratified data is included. However, as all these fits are from a model that is overparameterized for the length of time present in the data, it is difficult to find clear links between the datasets and the modelling output.

For the fits that incorporate data starting in 2000, the trends between the two fits are very similar but the fit with both datasets always has larger indices than the one with just fixed stations. Notably, the model estimates a low p_{nt} for the fixed stations, which implies that the abundance of non-target species is much larger than the number caught and results in lower capture rates (λ_{nt} , Figure 5). The one year where trends appear to differ between the two is in 2015, where the fit with both datasets has a spike noticeably absent from the fit with just the fixed stations. This is mostly caused by a couple sets that, due to the difference in p_{nt} and the large number of non-target species captured in those same sets, have a noticeably higher estimated halibut catch rate. Since these are located on the eastern edge of the area (see Figure 6), they end up being weighted heavily due to our spatially weighted average approach, which results in the spike in that given year which is not reflected as strongly in the random intercepts, which represent the expected catch rate after accounting for the vessel effects and the effect of space. Looking at the outcomes spatially for the fit to both datasets, the locations of high halibut catch rates appeared to be generally focused around the Scotian Shelf in the western part of the modelled area, with some spikes around the Laurentian Channel and a few isolated spikes on the eastern edge of the modelled area (Figure 6). For non-target catch rates, the highest spikes are consistently located on the western edge of the area towards the Gulf of Maine and the Bay of Fundy, with other less common spikes located east of the Laurentian Channel. The spatial outcomes from the fit with just the fixed stations has similar patterns (figures in Appendix 2).

3.4 Persistence

The persistence index of the spatial pattern for target species (halibut) indicated that the fixed stations are extremely likely (>98% with the highest persistence at ~ 0.1 , see table in Li et al. (2015)) to capture the abundance changes of this population abundance over time as the index is extremely low for all year pairings. However, the persistence of non-target species exhibits a different pattern (Figure 8, rescaled figure for target species available in Appendix B) wherein there appears to be a spatial shift in their distribution between early in the time series (mid-2000s) and the peak halibut abundance late in the time series (2017 to 2020). This would represent a higher chance that the fixed stations would miss changes in their distribution which, due to the nature of the multinomial model, would impact the estimation of target catch rates. There is some evidence of this happening in the fits with data from 2000 to 2021 in that the expected catch rate for non-target species increases in the late 2010s when the stratified data is included, but not when only the fixed stations are analyzed (Figure 3). However, these years also match up with a clear drop in the number of fixed stations (from ~ 150 to ~ 100). While the persistence analysis accounts for different numbers of stations between years, it is possible that the dropped stations were important for the distribution of non-target species, which should hopefully be compensated by the presence of the stratified random stations in the full analysis.

4 Discussion

This novel spatio-temporal MEM is able to successfully capture the changes of the halibut distribution and abundance over time and space for all available data with the exception of the first two years, including two different survey designs. It is the only model examined within

that is able to include all this information in a unified framework while extracting the maximum amount of information from the available data as possible. Furthermore, since this model is able to output both station-specific estimated catch rate and an overall estimated catch rate for the entire region, it both increases the amount of information available to manage this fishery while being directly comparable to current methods utilized to obtain an index of abundance.

The halibut stock assessment model requires indices of biomass instead of numbers of fish caught or catch rates [Trzcinski and Bowen (2016);Cox2018], meaning that the estimated catch rates from this model cannot be directly utilized by the assessment model. As the dataset contain average fish weight by set, it is a potential transformation away from having catch rates in biomass per hook per minute instead of in number of fish, which are then much more applicable and can still retain and propagate uncertainties forward through the delta method (Bickel and Doksum 2015). However, more work is necessary to be able to obtain the biomass output as stations that did not catch any halibut will still have an expected catch rate (albeit extremely low), but no fish weight associated with it, meaning that thought has to be put into how one would transform these low catch rates into biomass.

Similarly to previous versions of the MEM, our spatio-temporal MEM and our analysis of its performance on the Atlantic halibut fishery emphasizes the importance of accounting for the abundance of non-target species for a multivariate approach. The inability to estimate the probability of non-target escapes without hook condition information, illustrated very well in the fits that do not include the stratified dataset, demonstrates this impact very well. This underlines the importance of having a reliable estimate of the abundance of these non-target species. The modelling output themselves show that overestimating the abundance of these non-target species will likely result in an underestimation of the halibut population abundance. On the flip side, underestimating their abundance would likely result in the opposite, and completely ignoring them is likely to result in unpredictable and variable biases.

An interesting outcome of this analysis is that there is some evidence that, while the fixed stations were likely appropriate to capture the large changes of halibut abundance, they are likely not capturing changes in distribution and abundance of non-target species in recent years which would subsequently impact the estimation of halibut abundance. The persistence analysis on its own provides only relatively weak evidence due to having a high number of pairwise comparisons and because of the fact that the number of fixed stations dropped by a third in 2017, both which could be impacting the ability of the fixed stations to track the distribution of non-target species. However, because the intercepts obtained from the model when using only the fixed stations stay consistent but spike when using the stratified random stations, a change in the distribution of non-target species that would not have been captured by the fixed station design cannot be completely rejected. As the longline gear catches a great number of other species (e.g., cod, dogfish, etc.), it is not too surprising that their overall joint spatial distribution might shift which would then have an impact on the halibut catch rates. No matter if the true reason for the change is a shift in abundance or a drop in the number of fixed stations, the proportion of the catch that is non-target increased with the new stratified survey design in 2017. It is fortunate that with the implementation of this new design additional data on hook condition is collected such that this new model could be fit to more accurately estimate halibut abundance. Furthermore, the estimates from the stratified random design will only get more accurate and precise as years are added to the time series.

While the spatio-temporal MEM was developed for this specific longline survey, it could easily be applied to any longline fisheries or other passive sampling gear that would be impacted by hook or bait competition and promises to improve indices of abundance used for stock assessment models. For multi-species fisheries, the spatio-temporal MEM could be of great value to further modify the multinomial equation used by our approach to expand on the non-target species to include multiple specific species. More specifically, if there are a finite amount of other species being caught aside from the target, it would likely lead to improvements in estimated catch rates for the target species.

5 Acknowledgements

We want to thank the Atlantic Halibut Council for supporting this work, as well as survey coordinators, captains, crews and at-sea observers that have worked on the joint DFO-Industry survey without which none of this work would have been possible. We also thank Brendan Wringe for support on the original MEMSpa, and Jiaxin Luo for her creation of the original MEMSpa and for her help through code sharing and earlier exploration of the data. Finally, we would also like to thank both reviewers for helpful and constructive comments that helped improve this manuscript significantly.

6 References

- Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* 19(6): 716–723.
- Bickel, P.J., and Doksum, K.A. 2015. *Mathematical Statistics: Basic Ideas and Selected Topics*, Volumes I. Chapman; Hall/CRC Press.
- Cox, S., Benson, A., and Doherty, B. 2018. Re-design of the Joint Industry-DFO Atlantic Halibut (*Hippoglossus hippoglossus*) Survey off the Scotian Shelf and Grand Banks. (DFO Can. Sci. Advis. Sec. Res. Doc. 2018/020.): v + 50 p.
- den Heyer, C.E., Hubley, B., Themelis, D., Smith, S.C., Wilson, S., and Wilson, G. 2015. Atlantic Halibut on the Scotian Shelf and Southern Grand Banks : Data Review and Assessment Model Update. Canadian Science Advisory Secretariat Research Document DFO Can. S(2015/051).
- DFO. 2021. Stock Status Update of Atlantic Halibut (*Hippoglossus hippoglossus*) on the Scotian Shelf and Southern Grand Banks in NAFO Divisions 3NOPs4VWX5Zc for 2020. Canadian Science Advisory Secretariat (CSAS) (2021/024).
- Doherty, B., Cox, S., and Benson, A. (In press). March 2017 Technical Report March 2017 Technical Report. Landmark Fisheries.
- Etienne, M.-P., Obradovich, S., Yamanaka, L., and Mcallister, M. 2013. Extracting abundance indices from longline surveys : method to account for hook competition and unbaited hooks.: 1–35.

Kai, M., Thorson, J.T., Piner, K.R., and Maunder, M.N. 2017. Spatiotemporal variation in size-structured populations using fishery data: an application to shortfin mako (*Isurus oxyrinchus*) in the Pacific Ocean. *Canadian Journal of Fisheries and Aquatic Sciences* 74(11): 1765–1780.

Kess, T., Einfeldt, A.L., Wringe, B., Lehnert, S.J., Layton, K.K.S., McBride, M.C., Robert, D., Fisher, J., Le Bris, A., Den Heyer, C., Shackell, N., Ruzzante, D.E., Bentzen, P., and Bradbury, I.R. 2021. A putative structural variant and environmental variation associated with genomic divergence across the Northwest Atlantic in Atlantic Halibut. *ICES Journal of Marine Science* 78(7): 2371–2384.

Lee, L.M., and Rock, J.E. 2018. The forgotten need for spatial persistence in catch data from fixed-station surveys. *Fishery Bulletin* 116(1): 69–74.

Li, B., Cao, J., Chang, J.H., Wilson, C., and Chen, Y. 2015. Evaluation of Effectiveness of Fixed-Station Sampling for Monitoring American Lobster Settlement. *North American Journal of Fisheries Management* 35(5): 942–957.

Li, L., Hubley, B., Harper, D.L., Wilson, G., and Heyer, C.E. den. 2022. Data review and assessment model update: Assessment of atlantic halibut on the scotian shelf and southern grand banks (nafo divs. 3NOPs4VWX5Zc) data inputs and model. DFO Canadian Science Advisory Secretariat Research Document 2022/nnn: vi + xx p.

Li, Y., Lee, L.M., and Rock, J. 2019. Modeling population dynamics and nonstationary processes of difficult-to-age fishery species with a hierarchical bayesian two-stage model. *Canadian Journal of Fisheries and Aquatic Sciences* 76(12): 2199–2214.

Luo, J. 2020. Novel Statistical Analyses of Longline Survey Data for Improved Indices of Atlantic Halibut Abundance. Master's thesis, Dalhousie University.

Luo, J., McDonald, R.R., Wringe, B.F., den Heyer, C., Smith, B., Yan, Y., and Mills Flemming, J. 2022. A Spatial Analysis of Longline Survey Data for Improved Indices of Atlantic Halibut Abundance. Submitted to *ICES Journal of Marine Science*.

Pedersen, E.J., Goto, D., Gaeta, J.W., Hansen, G., Sass, G., Vander Zanden, M.J., Cichosz, T., and Rypel, A. 2018. Long-term growth trends in northern Wisconsin walleye populations under changing biotic and abiotic conditions. *Canadian Journal of Fisheries and Aquatic Sciences* 75(5): 733–745.

Rothschild, B.J. 1967. Competition for gear in a multiple-species fishery. *ICES Journal of Marine Science* 31(1): 102–110.

Schnute, J.T., and Richards, L.J. 1995. The influence of error on population estimates from catch-age models. *Canadian Journal of Fisheries and Aquatic Sciences* 52(10): 2063–2077.

Schwarz, G. 1978. Estimating the dimension of a model. *The Annals of Statistics* 6(2): 461–464.

Shackell, N.L., Fisher, J.A.D., den Heyer, C.E., Hennen, D.R., Seitz, A.C., Le Bris, A., Robert, D., Kersula, M.E., Cadrin, S.X., McBride, R.S., McGuire, C.H., Kess, T., Ransier, K.T., Liu, C., Czich, A., and Frank, K.T. 2021. Spatial Ecology of Atlantic Halibut across the Northwest Atlantic: A

Recovering Species in an Era of Climate Change. *Reviews in Fisheries Science and Aquaculture* 0(0): 1–25. Taylor & Francis.

Smith, S.J. 2016. Review of the Atlantic halibut longline survey index of exploitable biomass
Review of the Atlantic halibut longline survey index of exploitable biomass Stephen J . Smith
Science Branch , Maritimes Region Fisheries and Oceans Canada Bedford Institute of Oc.
Canadian Technical Report for Aquatic Sciences.

Swain, D.P., Jonsen, I.D., Simon, J.E., and Myers, R.A. 2009. Assessing threats to species at risk using stage-structured state - Space models: Mortality trends in skate populations. *Ecological Applications* 19(5): 1347–1364.

Trzcinski, M.K., Armsworthy, S.L., Wilson, S., Mohn, R.K., Fowler, M., and Campana, S.E. 2009. Atlantic Halibut on the Scotian Shelf and Southern Grand Banks (NAFO Divisions 3NOPs4VWX5ZC) - Industry/DFO Longline Survey and Tagging Results to 2008. Canadian Science Advisory Secretariat Science Advisory Report 026.

Trzcinski, M.K., and Bowen, W.D. 2016. The recovery of Atlantic halibut: a large, long-lived, and exploited marine predator. *ICES Journal of Marine Science* 73(4): 1104–1114.

Venables, W.N., and Dichmont, C.M. 2004. GLMs, GAMs and GLMMs: An overview of theory for applications in fisheries research. *Fisheries Research* 70: 319–337.

Warren, W. 1994. The potential of sampling with partial replacement for fisheries surveys. *ICES Journal of Marine Science* 51: 315–324.

Zwanenburg, K.C.T., and Wilson, S. 2000. The Scotian Shelf and Southern Grand Banks Atlantic halibut (*Hippoglossus hippoglossus*) survey — Collaboration between the fishing and fisheries science communities. *In* 2000/W:20. ICES CM.

Zwanenburg, K.C.T., Wilson, S., Branton, R., and Brien, P. 2003. Halibut on the Scotian Shelf and Southern Grand Banks - Current Estimates of Population Status. Canadian Science Advisory Secretariat (CSAS) (2003/046).

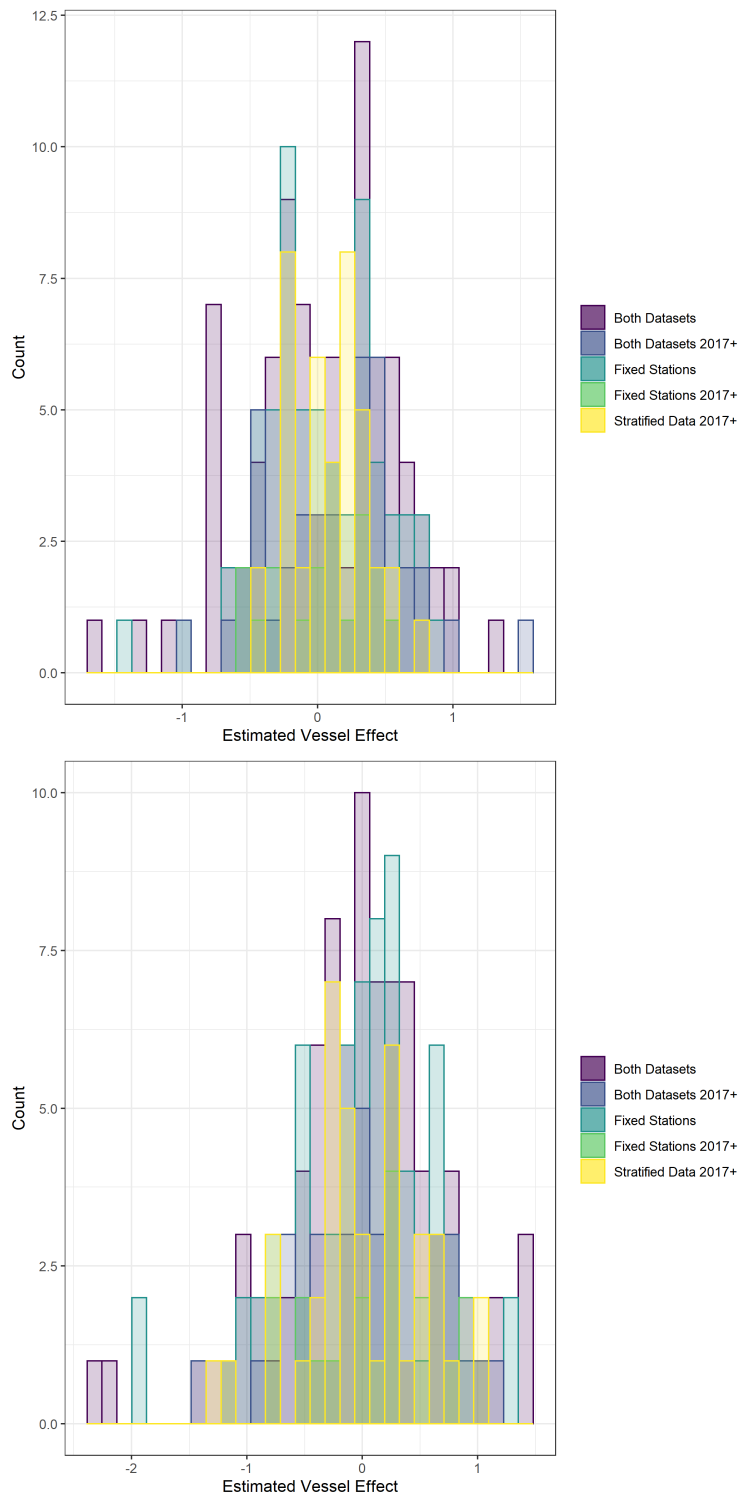


Figure 2. Distribution of estimated vessel effect using data from the fixed stations and stratified survey design both together and separately between 2000 and 2021 and between 2017 and 2021.

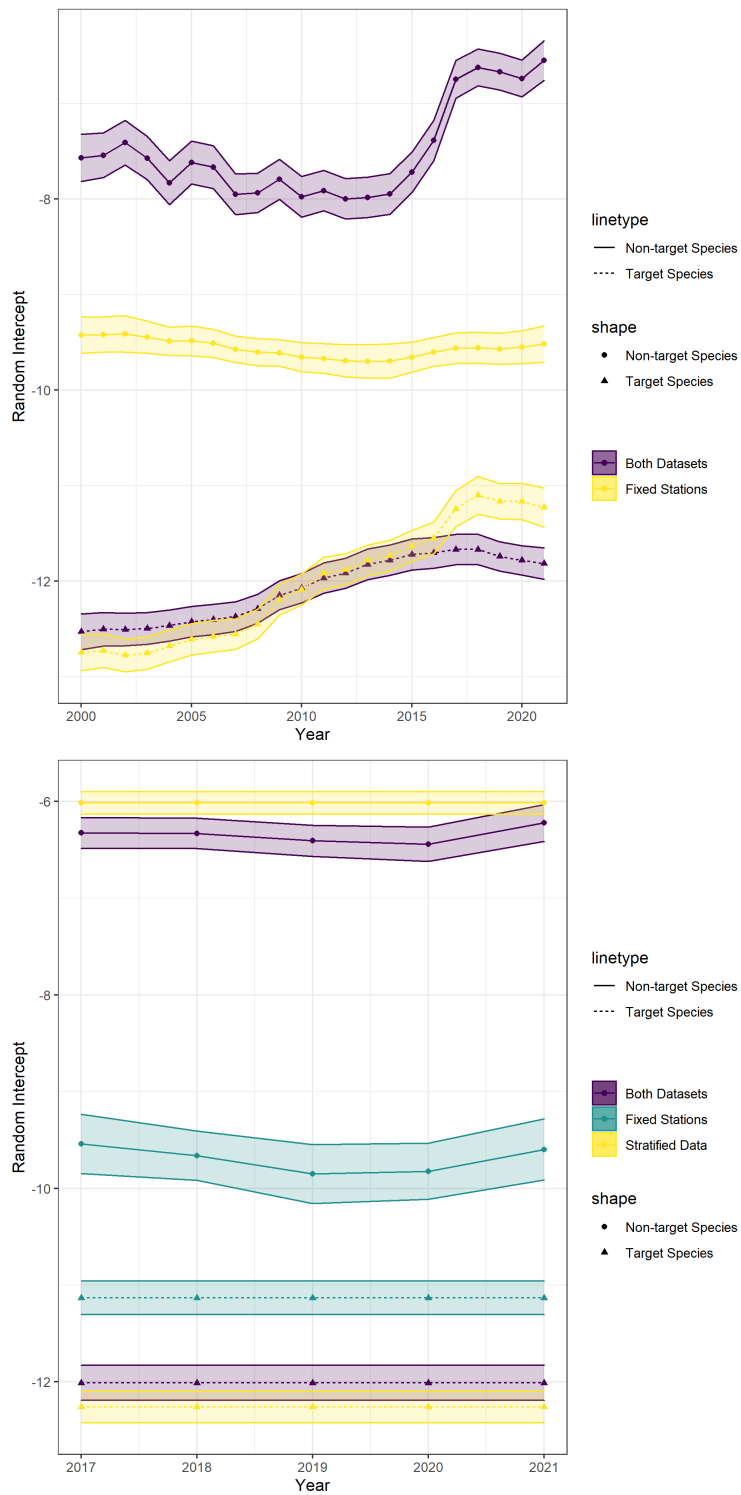


Figure 3. Estimated random intercepts using data from the fixed stations and stratified survey design both together and separately between 2000 and 2021 and between 2017 and 2021.

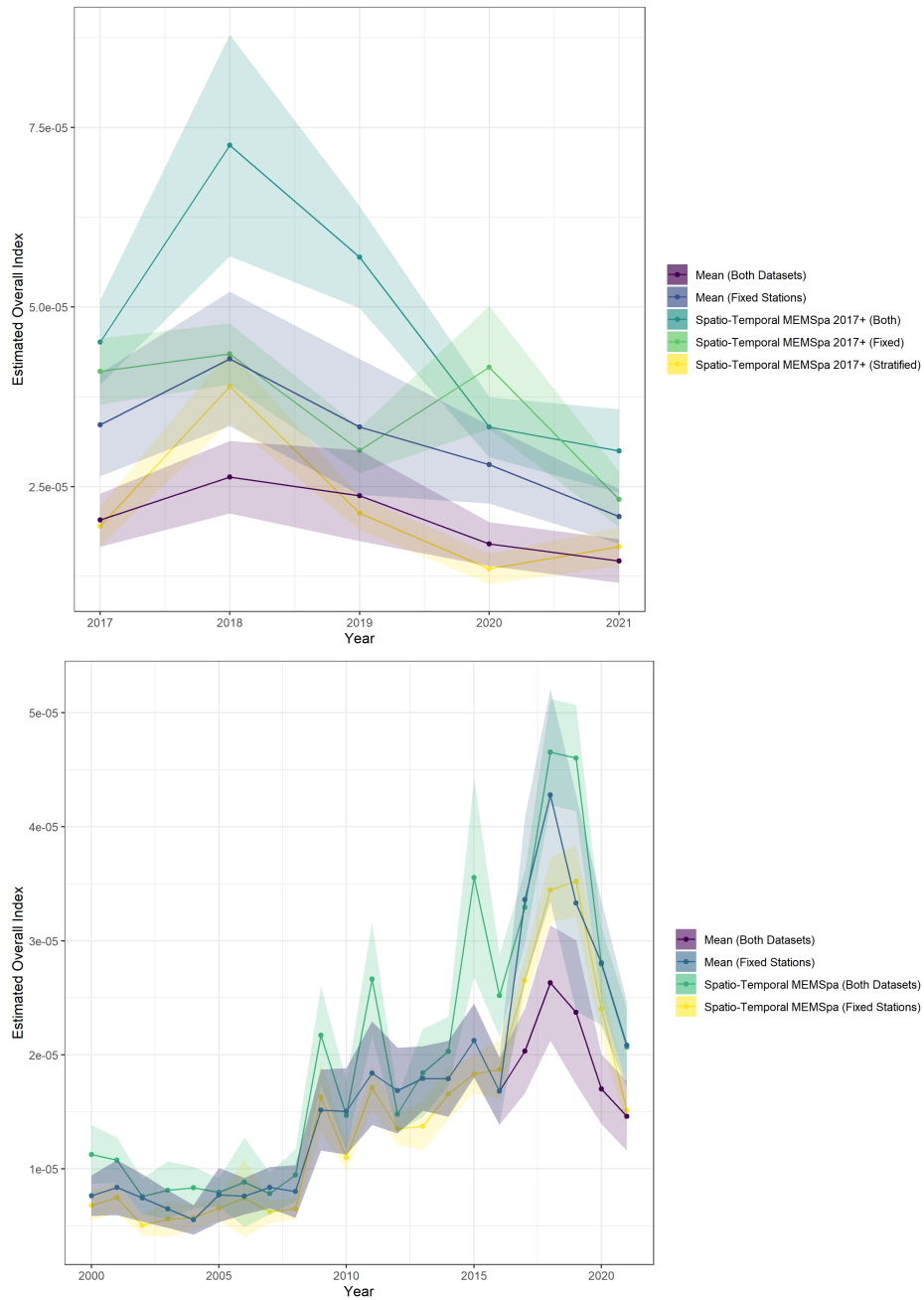


Figure 4. Estimated overall index for target species (expected number of halibut caught per hook per minute) using data from the fixed stations and stratified survey design both together and separately, the mean observed catch rate from the fixed stations and stratified survey design together, and the mean observed catch rate from the fixed stations from between 2000 and 2021 and between 2017 and 2021. The uncertainty represent +/- 1 standard error.



Figure 5. Estimated overall index for non-target species using data from the fixed stations and stratified survey design both together and separately between 2000 and 2021 and 2017 and 2021. The uncertainty around the model fits represent ± 1 standard error.



Figure 6. Station-specific estimated target species catch rates (number of animal per hook per minute) obtained using data from both the fixed stations and stratified survey design.

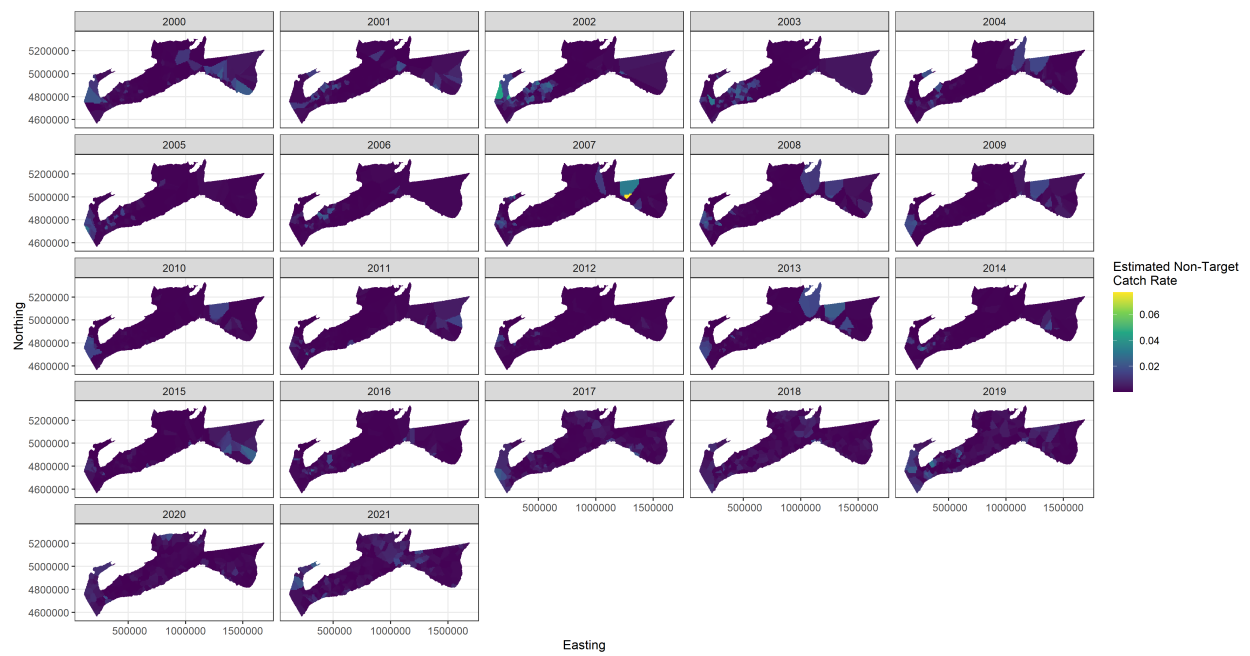


Figure 7. Station-specific estimated non-target species catch rates (number of animal per hook per minute) obtained using data from both the fixed stations and stratified survey design.

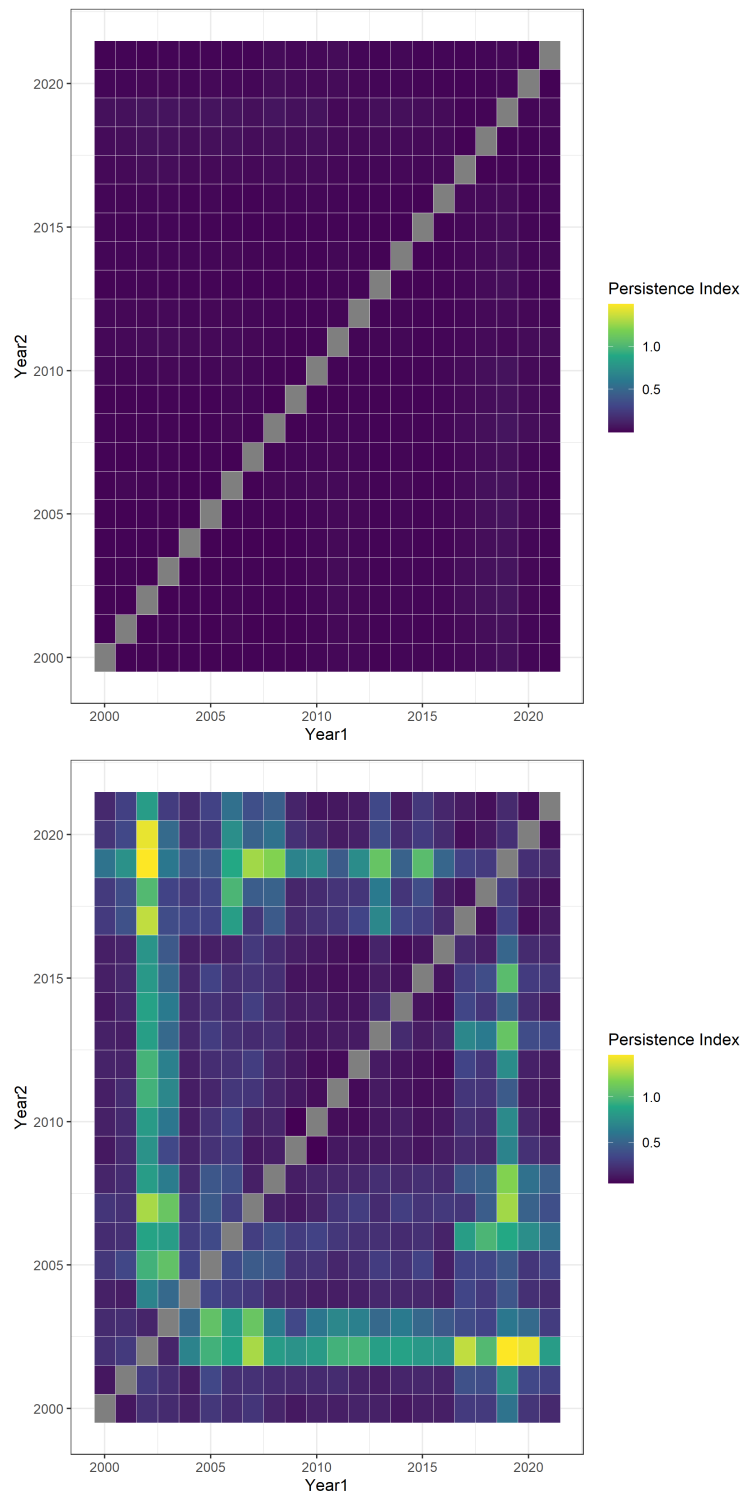


Figure 8. Persistence index between years for target species (top panel) and non-target species (bottom panel) catch rates.

APPENDIX A

2 September 2021

Selected Terms of Reference (TOR)

Statistical innovations for obtaining and validating Atlantic halibut longline survey indices of exploitable biomass through implementation of a multinomial, hook occupancy model.

Objectives of the Requirement

To extend the proposed spatial modeling to include the fix station and explore spatial temporal model to provide an index over time from the beginning of the survey until current time.

Background, Assumptions and Specific Scope of the Requirement

Together, the Atlantic Halibut Council and the Department of Fisheries and Oceans Canada (DFO) have used an annual longline survey to monitor Atlantic halibut exploitable biomass since 1988. The survey was originally stratified into areas of Low, Medium and High catch based on data from commercial fishing logs (1995-1997). Stratified estimates were used until the assessment by Trzcinski et al. (2009) when the stratification system was no longer utilized (although the strata were still part of the survey design). Starting in 2009, a standardized catch rate calculated from a negative binomial (NB) generalized linear model (GLM) replaced the stratified estimate of mean weight per standard longline set (den Heyer et. al., 2013).

The simple stratified mean adjusted catch rate and the NB GLM both assume that halibut are the only species being caught by the longline hooks with no accounting for other species competing for hooks. In addition, these methods implicitly assume that all hooks that do not have halibut would still have been able to catch halibut had there been more of these fish in the area. However, some number of hooks will be occupied by species other than halibut and other hooks will be empty with bait still attached or missing. Smith (2016) recently recommended replacing the NB GLM with a multinomial model that accounted for both the number of halibut caught, and the number of hooks occupied by other species or missing bait.

The Atlantic halibut longline survey was designed to provide an annual index of abundance (numbers or weights) to monitor the status of the stock for management purposes. Based on the work has been done by now, it is useful to extend that spatial modeling to include the fix station as well, also explore spatial temporal model to provide an index over time from the beginning of the survey until current time. In other words, we will have one index to include in our assessment model for the longline survey data both from fixed station and stratified random.

Tasks, Activities and Deliverables

The deliverables for the project will be to:

- 1) Develop spatiotemporal multinomial model for stratified random station survey 2017-2020
- 2) Develop spatiotemporal multinomial model for the fixed station survey data 1998-2020 (app 230 station 1998-2016 and 100 stations 2017-2020)

- 3) Develop a spatiotemporal model that simultaneously analyzes the data from the four years, 2017-2020, for which there are both fixed and stratified random station data.
- 4) Develop a spatiotemporal model for the entire time series, 1998-2020, of fixed station and random survey data
- 5) A technical report.

References

- Armstrong, S, Wilson, S, Mohn, R. 2006. Atlantic halibut on the Scotian Shelf and Southern Grand Banks (NAFO Division 3NOPs4VWX5Zc) – Industry/DFO longline survey results to 2005. DFO Can. Sci. Advis. Sec. Res. Doc. 2006/065. ii +31p.
- den Heyer, C, Schwarz, C, and Trzcinski, M. 2013. Fishing and natural mortality rates of Atlantic halibut estimated from multiyear tagging and life history. *Trans. Am. Fish. Soc.* 143(3): 690-702.
- Etienne, M, Obradovich, S, Yamanaka, K, and McAllister, M. 2013. Extracting abundance indices from longline surveys: a method to account for hook competition and unbaited hooks. *arXiv* 1005.0892v3:1-35.
- Smith, S. 2016. Review of the Atlantic halibut longline survey index of exploitable biomass. Canadian Technical Report of Fisheries and Aquatic Sciences 3180.
- Thorson, J, Ianelli, J, Munch, S, Ono, K, Spencer, P. 2015b. Spatial delay-difference models for estimating spatiotemporal variation in juvenile production and production abundance. *Can. J. Fish. Aquat. Sci.* 72 (12), 1897-1915. doi: 10.1139/cjfas-2014-0543.
- Thorson, J, Skaug, H, Kristensen, K, Shelton, A, Ward, E, Harms, J, Benante, J. 2015a. The importance of spatial models for estimating the strength of density dependence. *Ecology* 96 (5), 1202–1212. doi: 10.1890/14-0739.1.sm
- Thorson, J, Shelton, A, Ward, E., Skaug, H. 2015c. Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. *ICES J. Mar. Sci.* 72, 1-11. doi: 10.1093/icesjms/fst176.
- Trzcinski, M, Armstrong, S, Wilson, S, Mohn, R, Fowler, M and Campana, S. 2009. Atlantic halibut on the Scotian Shelf and Southern Grand Banks (NAFO Division 3NOPs4VWX5Zc) – Industry/DFO longline survey and tagging results to 2008. DFO Canadian Scientific Advisory Sec. Res. Doc. 2009/026. vi +43p

APPENDIX B

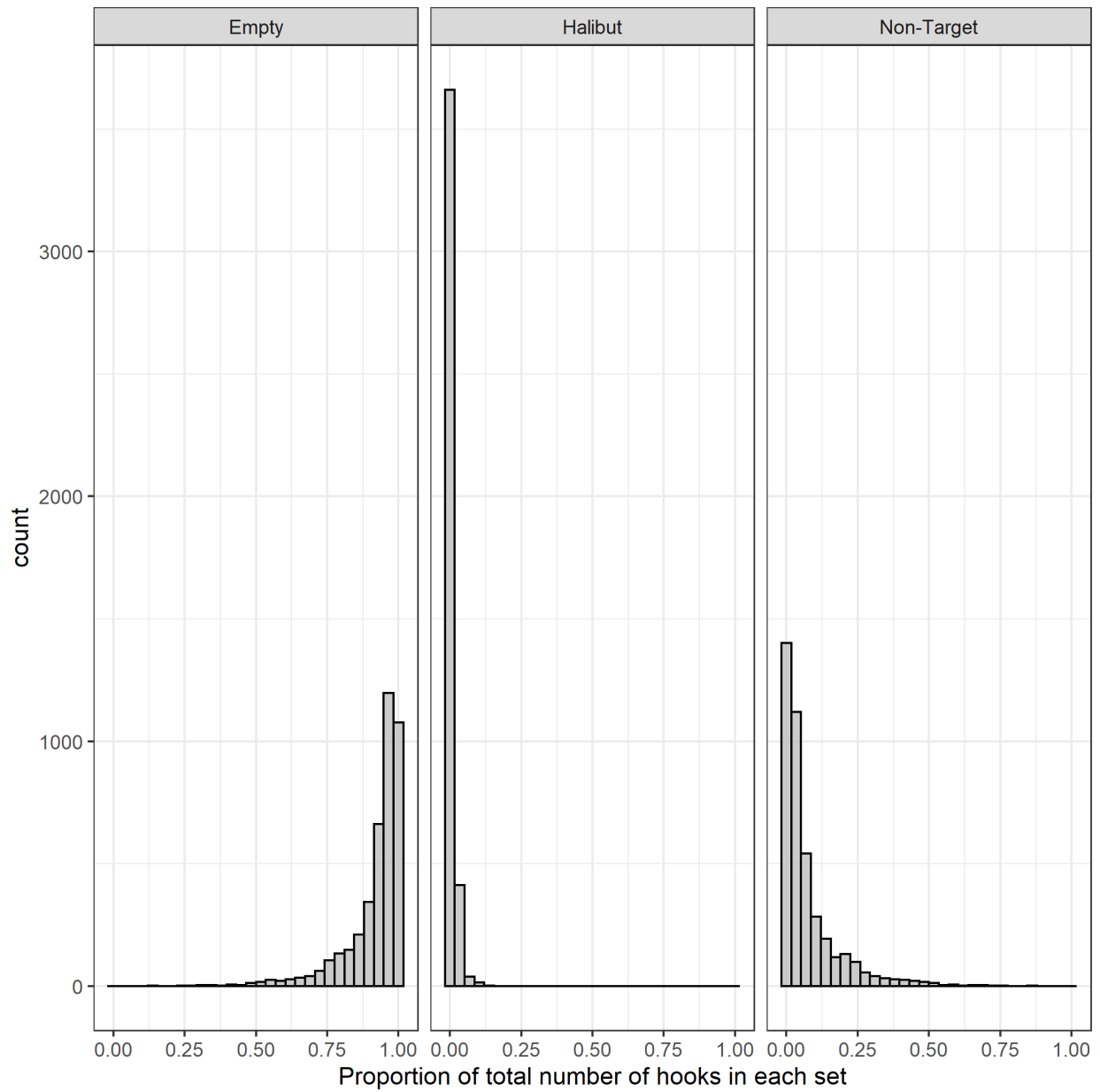


Figure B.1. Histogram of proportion of each category in each set for fixed stations data.

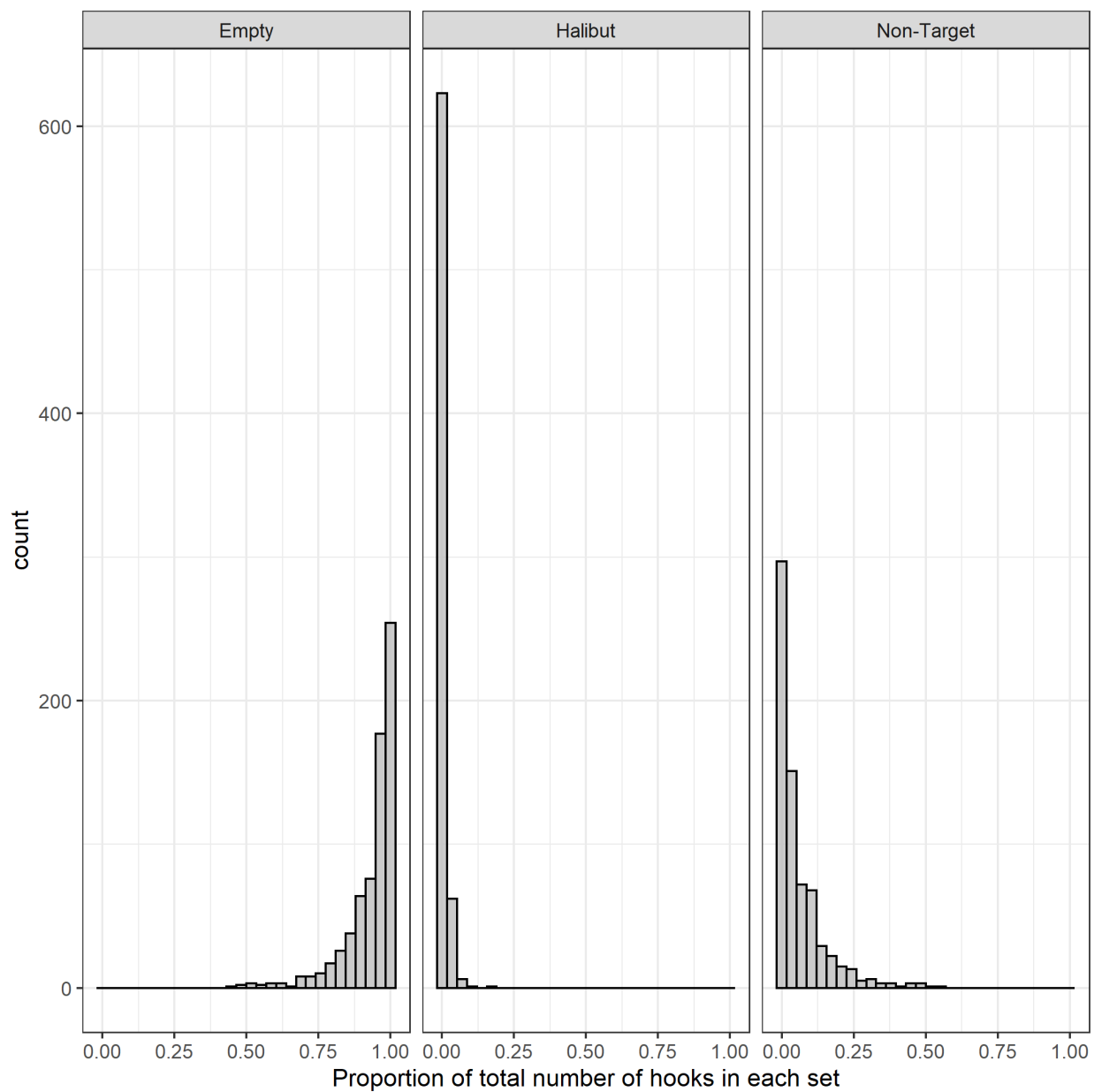


Figure B.2. Histogram of proportion of each category in each set for stratified dataset.

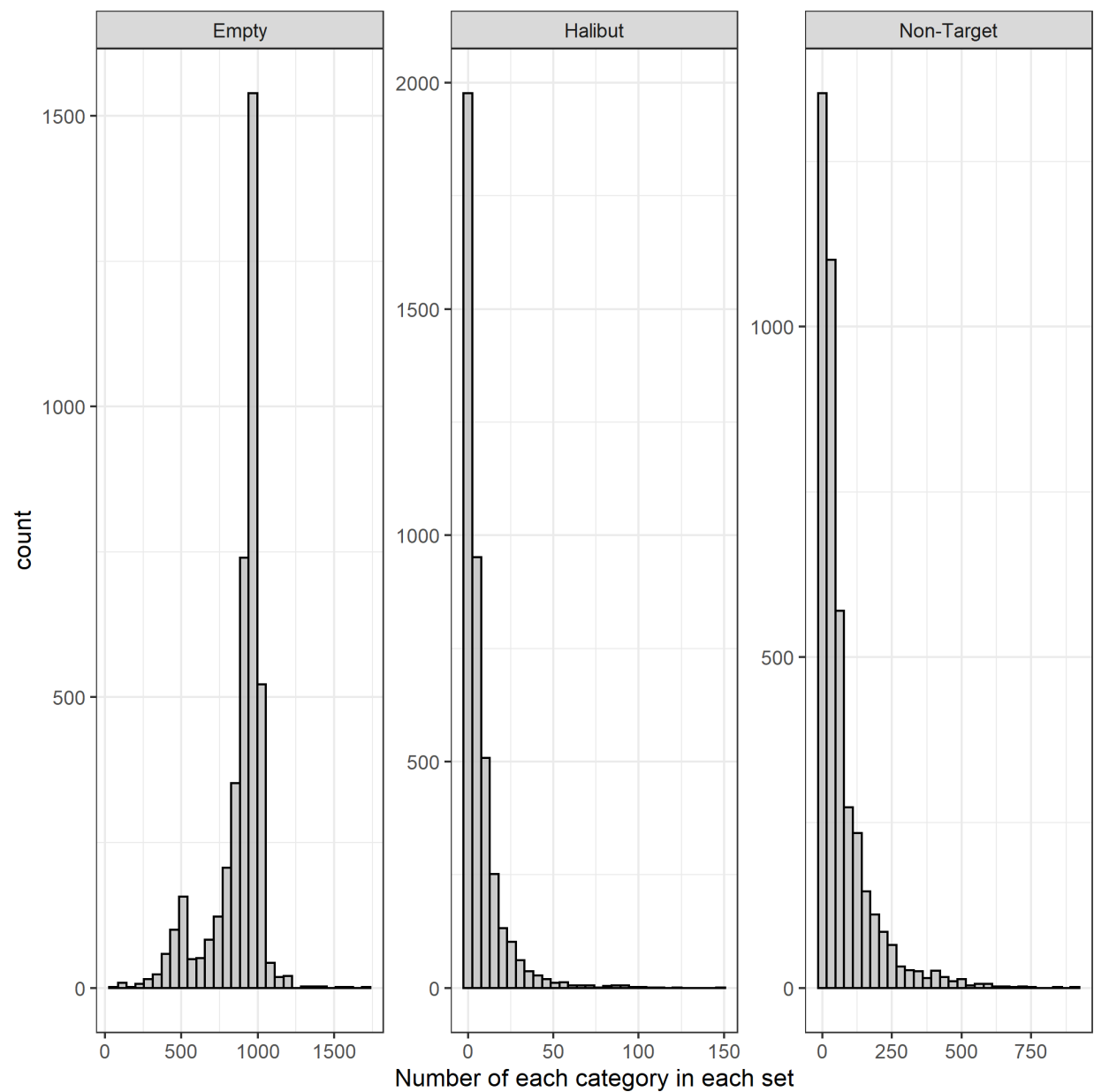


Figure B.3. Histogram of numbers of each category in each set for fixed stations data.

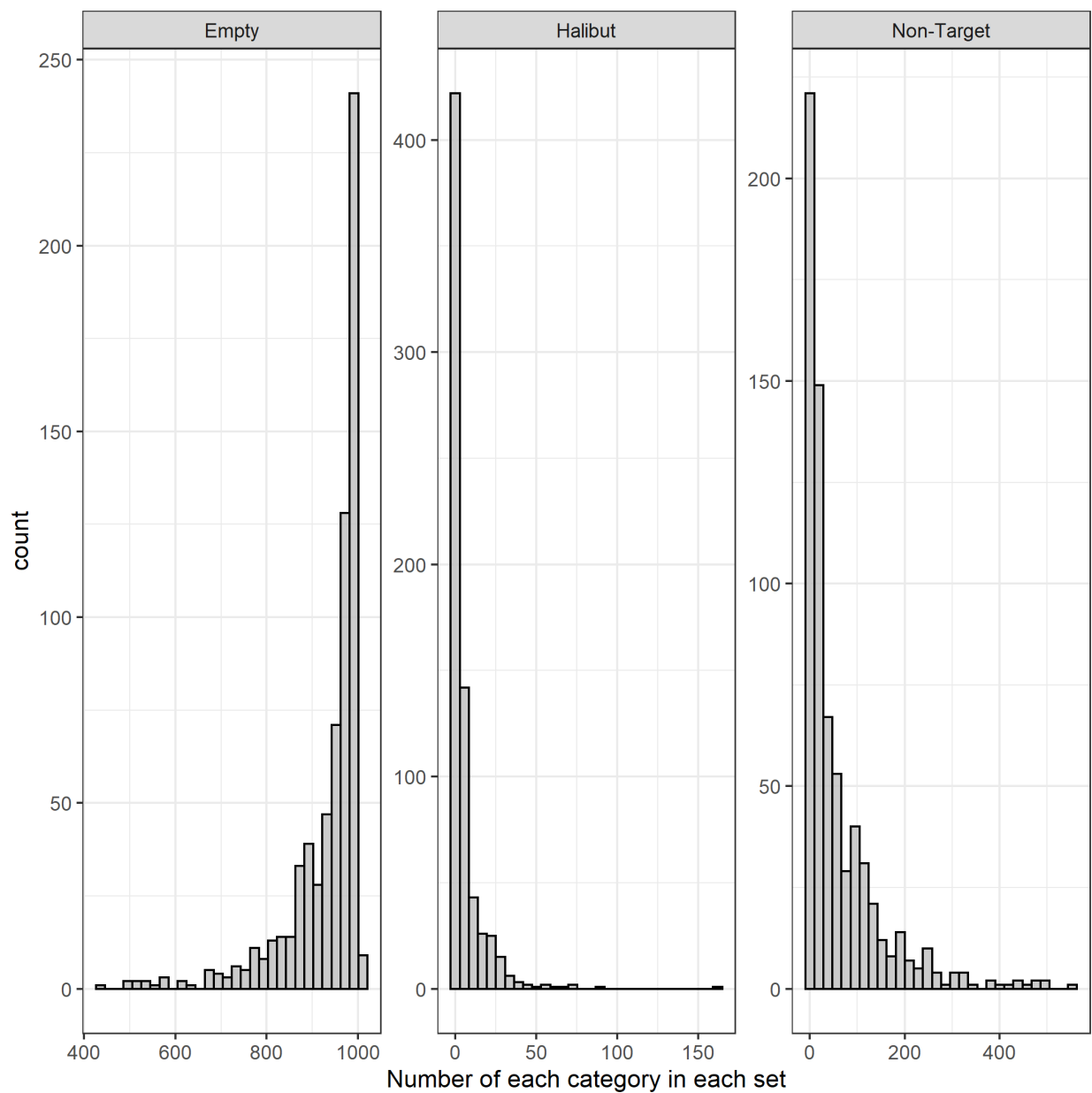


Figure B.4. Histogram of numbers of each category in each set for stratified dataset.



Figure B.5. Station-specific estimated target species catch rates obtained using data from only the fixed stations.

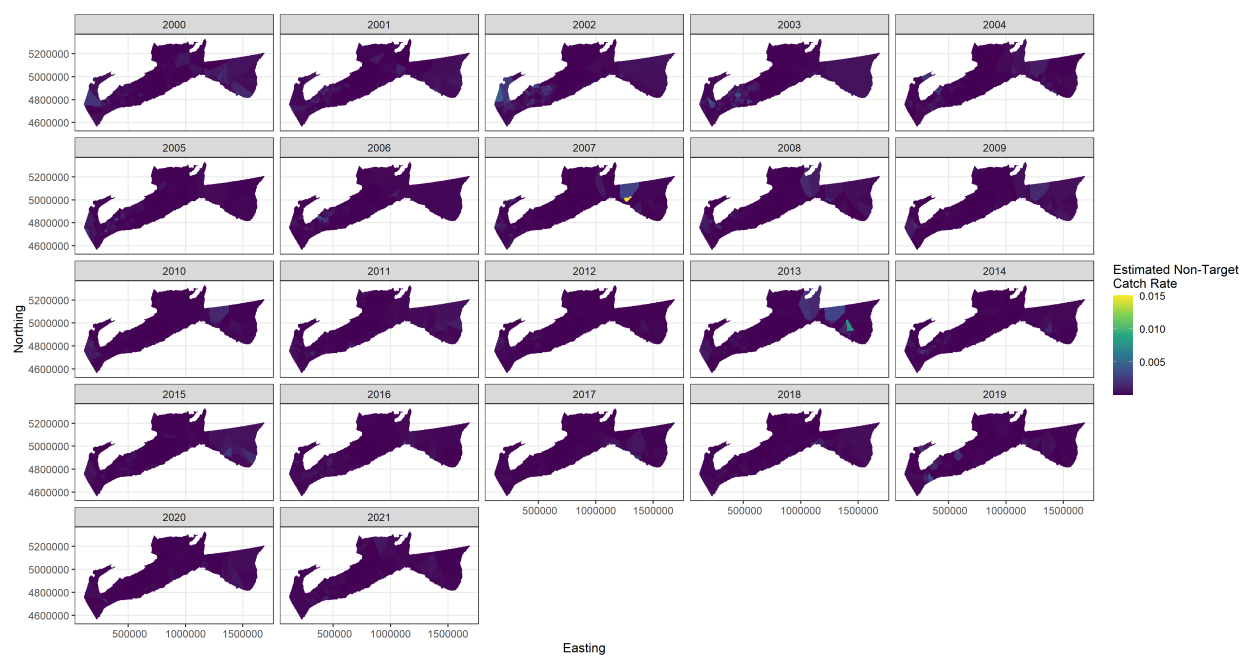


Figure B.6. Station-specific estimated non-target species catch rates obtained using data from only the fixed stations.

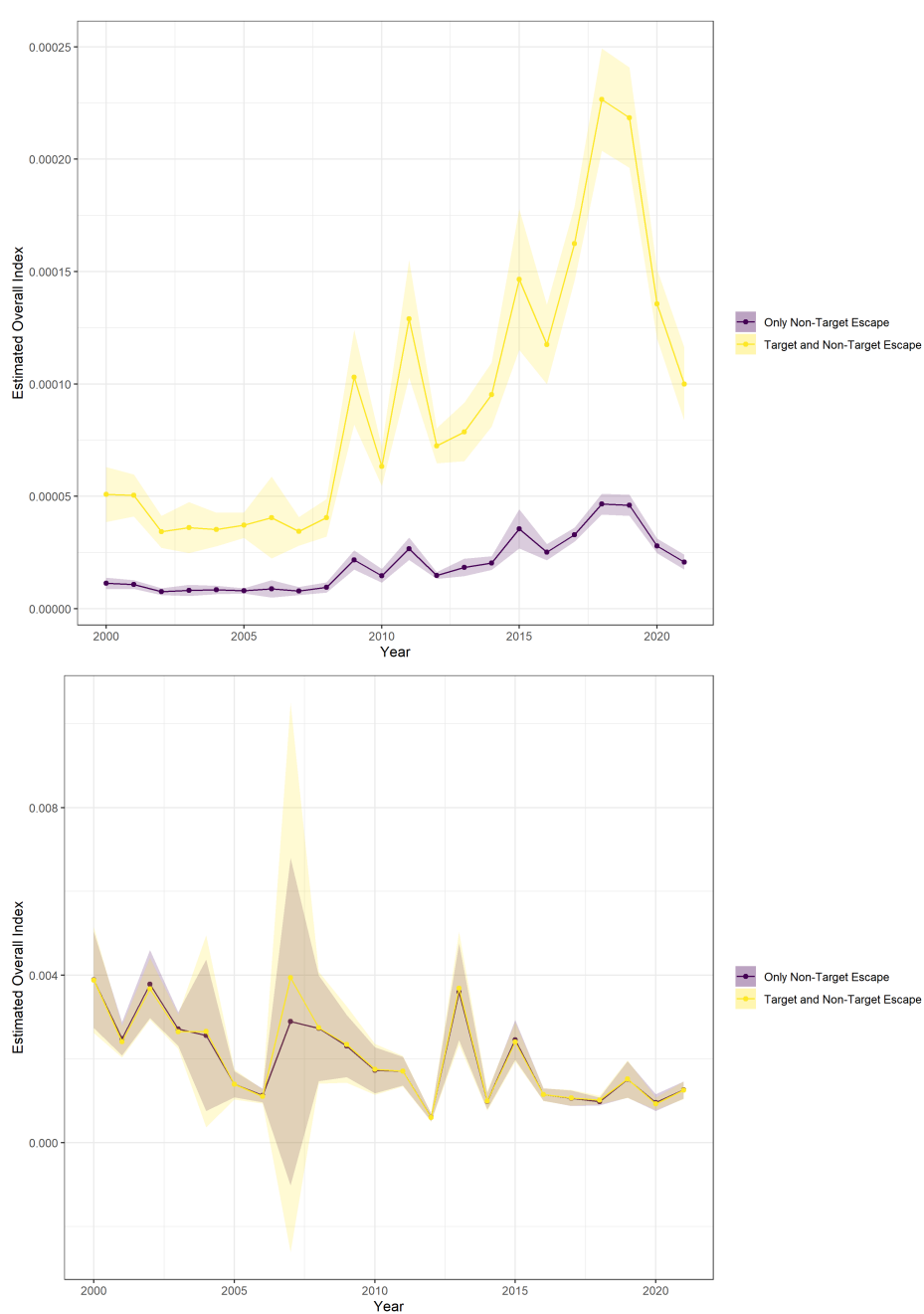


Figure B.7. Estimated overall index for target and non-target species (expected number of halibut caught per hook per minute) by the random walk model where only non-target species can escape and the same model where target species and non-target species have an identical probability of escape using data from the fixed stations and stratified survey design together between 2000 and 2021 and between 2017 and 2021. The uncertainty represent +/- 1 standard error.

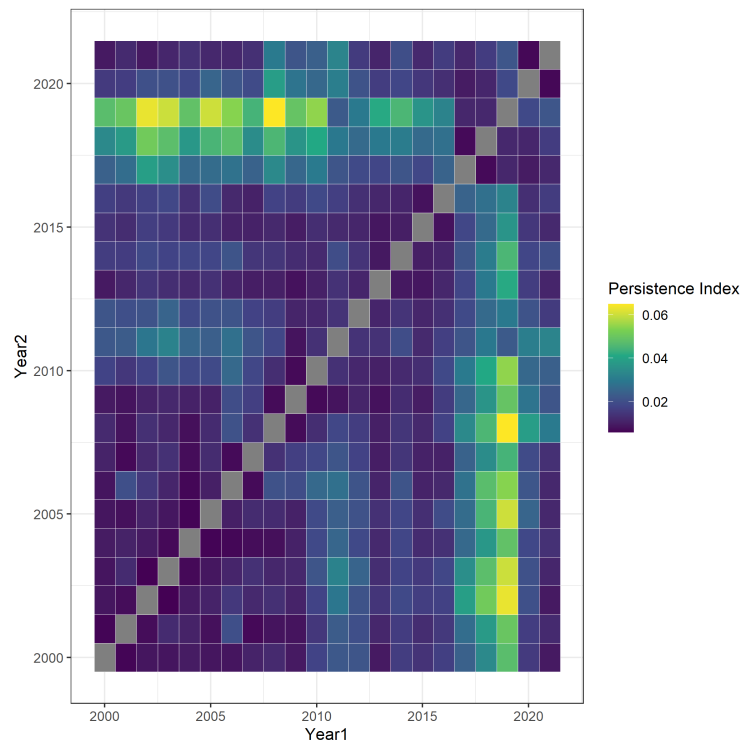


Figure B.8. Persistence index between years for target species catch rates.

APPENDIX C

Table C.1. Outputs for model selection approaches when fitting various models to both datasets from 2017 to 2021, including root mean squared errors (RMSE), Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC).

Model	Observation RMSE	Target Field RMSE	Non-Target Field RMSE	AIC	BIC
Random Intercept	0.7678375	1.587530	1.381389	108,129.9	108,185.8
Random Walk	0.7678405	1.587320	1.375622	108,181.7	108,237.5
Random Slope	0.7678375	1.587530	1.381383	108,129.9	108,185.8
AR(1) Process	0.7678409	1.584584	1.374446	108,189.7	108,255.8

Table C.2. Outputs for model selection approaches when fitting various models to the fixed stations dataset from 2000 to 2021, including root mean squared errors (RMSE), Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC).

Model	Observation RMSE	Target Field RMSE	Non-Target Field RMSE	AIC	BIC
Random Intercept	0.0017602	1.187879	1.463658	56,782.04	56,851.63
Random Walk	0.0017645	1.190783	1.469420	56,698.79	56,768.37
Random Slope	0.0017602	1.187879	1.463655	56,782.01	56,851.59
AR(1) Process	0.0017605	1.187945	1.463765	56,785.64	56,867.87

Table C.3. Outputs for model selection approaches when fitting various models to the fixed stations dataset from 2017 to 2021, including root mean squared errors (RMSE), Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC).

Model	Observation RMSE	Target Field RMSE	Non-Target Field RMSE	AIC	BIC
Random Intercept	0.0012898	1.168684	1.431670	7,459.491	7,505.696
Random Walk	0.0012912	1.168679	1.434893	7,449.267	7,495.472
Random Slope	0.0012898	1.168684	1.431671	7,459.483	7,505.688
AR(1) Process	0.0012882	1.167309	1.433463	7,455.844	7,510.450

Table C.4. Outputs for model selection approaches when fitting various models to the stratified stations dataset from 2017 to 2021, including root mean squared errors (RMSE), Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC).

Model	Observation RMSE	Target Field RMSE	Non-Target Field RMSE	AIC	BIC
Random Intercept	0.7932118	1.556453	0.8235129	84,070.09	84,120.04
Random Walk	0.7932118	1.556453	0.8235128	84,065.74	84,115.70
Random Slope	0.7932118	1.556453	0.8235036	84,076.36	84,126.31
AR(1) Process	0.7932122	1.557215	0.8223264	84,079.74	84,138.77

Table C.5. RMSE obtained from 10-fold cross-validation output for each fold using the random walk model to both datasets from 2000 to 2021

CV Fold	Overall RMSE	Target RMSE	Other RMSE
1	44.70686	6.630367	60.38271
2	43.65656	6.211522	57.48138
3	41.96582	4.933174	54.23545
4	42.77388	5.864730	56.85416
5	48.66753	5.720658	64.42231
6	43.31645	4.954222	58.60077
7	39.43232	4.880780	50.27927
8	45.40232	5.721324	59.55889
9	40.26005	4.799672	53.77724
10	41.67242	5.932196	53.49398