

NOVEL STATISTICAL ANALYSES OF LONGLINE SURVEY
DATA FOR IMPROVED INDICES OF ATLANTIC HALIBUT
ABUNDANCE

by

Jiixin Luo

Submitted in partial fulfillment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
December 2020

© Copyright by Jiixin Luo, 2020

Dedicated to myself, my parents and everyone else who made this possible. Thanks a million!

Table of Contents

List of Tables	v
List of Figures	vii
Abstract	x
Acknowledgements	xi
Chapter 1 Introduction	1
1.1 Background	1
1.2 Data Description	4
1.3 Survey Index Reproduction	19
1.3.1 Multinomial Exponential Model	19
1.3.2 Template Model Builder (TMB)	22
1.3.3 Comparison of Results	23
Chapter 2 Spatial Models and Variable Selection	26
2.1 Spatial Patterns	26
2.2 Models with a Gaussian Random Field	29
2.2.1 Binomial Model with Gaussian Random Field (glmmTMB)	29
2.2.2 Multinomial Exponential Model (MEM1) with Gaussian Random Field (TMB)	32
2.2.3 Variable Selection	35
Chapter 3 Estimated Survey Indices	40
3.1 Estimated Survey Indices with Corresponding Standard Errors	40
3.2 Interpolation of Estimated Relative Abundance Indices and Random Field	44
3.3 Weighted Average of the Survey Indices	48
3.3.1 Explicit Method and corresponding Standard Error	48
3.3.2 Mean of the Interpolated Survey Indices and the Corresponding Standard Error	48

3.3.3	Dirichlet Method and the Corresponding Standard Error . . .	49
Chapter 4	Discussion	52
Bibliography	55
Appendix	58
A.1	Survey Indices Reproduction	58
A.2	Binomial Model with Gaussian Random Field (glmmTMB)	62
A.3	Multinomial Exponential Model (MEM1) with Gaussian Random Field without Covariate (TMB)	65
A.4	Interpolation of Estimated Relative Abundance Indices and Random Field	74
A.5	Aggregation with Dirichlet Method	77

List of Tables

1.1	Partial longline survey data for 2017.	6
1.2	The planned (in brackets) and completed allocation of survey stations per stratum and the total number of survey stations.	8
1.3	Summary of the number of stations with and without observed temperature data and the percentage of stations with observed temperature.	13
1.4	The estimated relative abundance indices across the entire survey area with mean of the soak time for the survey for target and non-target species using R packages TMB and llsurv for hook occupancy data from the halibut longline survey for the years 2017, 2018 and 2019.	24
1.5	The estimated relative abundance indices across the entire survey area with set-specific soak time for target and non-target species and the escaping probability of caught non-target species using the TMB package for hook occupancy data from the halibut longline survey for the years 2017, 2018 and 2019.	24
2.1	Summary of the ANOVA test of proportion of unbaited hooks by strata for 2017.	28
2.2	Summary of the ANOVA test of proportion of unbaited hooks by strata for 2018.	28
2.3	Summary of the ANOVA test of proportion of unbaited hooks by strata for 2019.	28
2.4	AIC and BIC for different models for 2017.	36
2.5	AIC and BIC for different models for 2018.	37
2.6	AIC and BIC for different models for 2019.	38
3.1	Estimated parameters of Matérn covariance function for target species for 2017, 2018 and 2019. Smoothness parameter ν is fixed at one.	45

3.2	Estimated parameters of Matérn covariance function for non-target species for 2017, 2018 and 2019. Smoothness parameter ν is fixed at one.	45
-----	--	----

List of Figures

1.1	Photograph of an Atlantic Halibut [2].	1
1.2	The survey area for the 3NOPs4VWX5Zc Atlantic Halibut stock (with five area strata: 4X5YZ, 4W, 4V, 3P and 3NO).	4
1.3	Locations of the 150 survey stations in 2017.	7
1.4	Locations of the 153 survey stations in 2018.	7
1.5	Locations of the 153 survey stations in 2019.	8
1.6	Absolute differences (m) between survey depth and depth obtained using the bathymetric method for 2017.	10
1.7	Absolute differences (m) between survey depth and depth obtained using the bathymetric method for 2018 without stations 570, 572 and 573.	10
1.8	Absolute differences (m) between survey depth and depth obtained using the bathymetric method for 2019 without stations 570, 571, 572, 573 and 574.	11
1.9	Locations of the survey stations with indication of absolute difference between survey depth and bathymetric depth for 2017.	11
1.10	Locations of the survey stations with indication of absolute difference between survey depth and bathymetric depth for 2018.	12
1.11	Locations of the survey stations with indication of absolute difference between survey depth and bathymetric depth for 2019.	12
1.12	Observed temperature and model estimated temperature for 2017 and 2018.	14
1.13	The residuals between observed temperature and model estimated temperature for 2017 and 2018.	14
1.14	The residuals between observed temperature and model estimated temperature for 2017.	15
1.15	The residuals between observed temperature and model estimated temperature for 2018.	15

1.16	Histograms of the soak time by year for 2017 to 2019.	16
1.17	Histograms of the soak time by depth strata for 2017 to 2019.	17
1.18	Histograms of the soak time by area for 2017 to 2019.	18
1.19	A comparison of the proportion of halibut and non-target species caught with different soak times for 2017, 2018 and 2019.	25
2.1	The distribution of unbaited hooks in 2017.	27
2.2	The distribution of unbaited hooks in 2018.	27
2.3	The distribution of unbaited hooks in 2019.	28
2.4	The predicted probabilities of getting unbaited hooks at 1000 randomly selected locations for 2017.	31
2.5	The predicted probabilities of getting unbaited hooks at 1000 randomly selected locations for 2018.	31
2.6	The predicted probabilities of getting unbaited hooks at 1000 randomly selected locations for 2019.	32
2.7	The correlation matrix for both target and non-target species for 2017.	33
2.8	The correlation matrix for target species for 2017.	34
2.9	The correlation matrix for non-target species for 2017.	34
2.10	Relative abundance indices for halibut and corresponding standard errors estimated by models with and without temperature for 2017.	39
2.11	Relative abundance indices for halibut and corresponding standard errors estimated by models with and without temperature for 2018.	39
3.1	Estimated relative abundance indices for halibut for 2017 (top), 2018 (middle) and 2019 (bottom) based on the multinomial exponential model with Gaussian random field.	41
3.2	Boxplots of classified estimated relative abundance indices based on the multinomial exponential model with Gaussian random field for 2017, 2018 and 2019.	42

3.3	Estimated relative abundance indices and corresponding standard errors for halibut for 2017 based on the multinomial exponential model with Gaussian random field.	43
3.4	Estimated relative abundance indices and corresponding standard errors for halibut for 2018 based on the multinomial exponential model with Gaussian random field.	43
3.5	Estimated relative abundance indices and corresponding standard errors for halibut for 2019 based on the multinomial exponential model with Gaussian random field.	44
3.6	Interpolated Gaussian random fields for halibut for 2017 (top), 2018 (middle) and 2019 (bottom).	46
3.7	The heatmaps of the interpolated relative abundance indices of halibut for 2017 (top), 2018 (middle) and 2019 (bottom). . . .	47
3.8	The Dirichlet tessellation and the survey stations for 2017 (top), 2018 (middle) and 2019 (bottom).	50
3.9	The weighted survey indices and corresponding standard errors for halibut obtained using three different weighting methods (all strata except H13) for 2017, 2018 and 2019.	51
4.1	Relative abundance indices for halibut and corresponding standard errors estimated by models with set-specific soak time and mean soak time for 2017.	53
4.2	Relative abundance indices for halibut and corresponding standard errors estimated by models with set-specific soak time and mean soak time for 2018.	53
4.3	Relative abundance indices for halibut and corresponding standard errors estimated by models with set-specific soak time and mean soak time for 2019.	54

Abstract

The longline halibut survey is a joint project between the Atlantic Halibut Council (AHC) and Fisheries and Oceans Canada (DFO). This annual survey is used to monitor the status of Atlantic Halibut (*Hippoglossus hippoglossus*) on the Scotian Shelf and southern Grand Banks (NAFO Divisions 3NOPs4VWX5Zc). The purpose of this thesis is to develop novel statistical analyses of Atlantic Halibut longline survey data for more accurate and precise indices of relative abundance. Reproduction of the survey indices has been completed using hook occupancy data collected between 2017 and 2019 and current hook competition model. In order to reflect spatial patterns in these data, a spatial random field is introduced into the model. Covariates for depth, temperature and area strata (along with their potential interactions) are also introduced as these may affect the survey indices. Model selection has been performed and the selected spatial model is then used to estimate the relative abundance indices for target and non-target species at each survey station for 2017 to 2019. Finally, the estimated indices are aggregated over space using three different approaches to compare the estimated changes in survey indices with those obtained with the non-spatial model currently being used by DFO. The results indicate that the relative abundance indices for halibut are higher on shallow banks in southwest Nova Scotia (NAFO 4X and 4W) and along the shelf edge throughout the management unit.

Acknowledgements

I would like to express the deepest appreciation to my supervisors Bruce Smith and Joanna Mills Flemming for their persistent support. They guided me through the thesis research, contributed in countless ways to my development and encouraged me all the time.

A special thanks to Brendan Wringe and Nell den Heyer for providing patient and timely responses to all of my questions about the database and the survey. I would also like to thank Zeliang Wang for providing the estimated bottom temperature data.

I would like to thank Yuan (Sofie) Yan for answering all my questions about Spatial Statistics with patience and Ethan Lawler for helping me code with R package TMB.

Last but not the least, I would like to acknowledge the rest of my thesis committee for spending time reading my thesis and providing valuable comments on this thesis.

Chapter 1

Introduction

We introduce the annual Atlantic Halibut longline survey and describe the survey data in the first two sections of this chapter. We then discuss the models and R packages that are currently used to obtain the survey indices.

1.1 Background

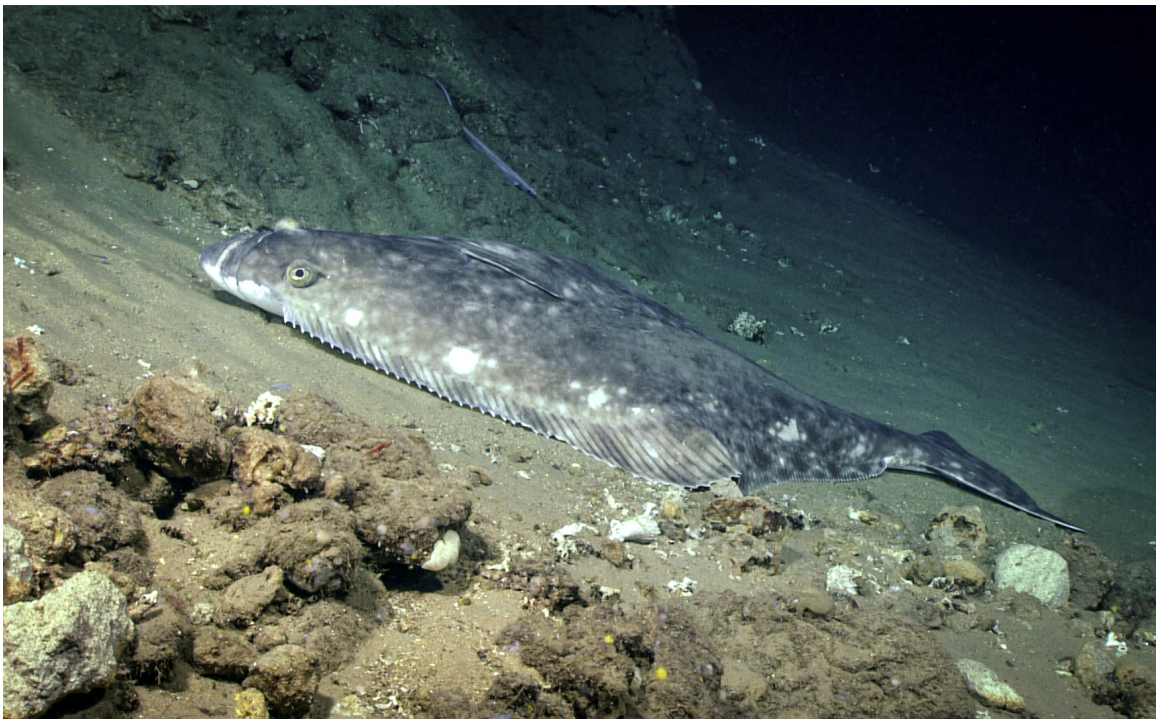


Figure 1.1: Photograph of an Atlantic Halibut [2].

Atlantic Halibut is an economically important demersal flatfish species that is distributed widely throughout the North Atlantic Ocean, as well as being found in parts of the Arctic Ocean. It is commonly found in deep-water channels running between banks and along the edge of the continental shelf. The depth distribution of Atlantic

Halibut ranges from less than 50 m to over 1000 m but they are most abundant in waters from 200 m to 500 m deep [1]. There are two Atlantic Halibut management units in Canadian waters, namely the Scotian Shelf and southern Grand Banks (North Atlantic Fisheries Organization [NAFO] Divisions 3NOPs4VWX5Zc), and the Gulf of St. Lawrence (4RST). This analysis will focus on the 3NOPs4VWX5Zc stock (Figure 1.2).

Together, the AHC and DFO have used an annual longline survey to monitor Atlantic Halibut exploitable biomass since 1998. Longline fishing is a commercial fishing technique that enables massive fish yields and the length of the longline for halibut survey is approximately 5 km. Longlines can be placed at the sea surface (pelagic longlines) or at the seafloor (demersal longlines) and consist of a very long mainline. Baited hooks, used to lure and capture target species, are attached to the mainline by branchlines. The survey is completed by commercial fishermen with at-sea on board observers collecting data. The survey was originally stratified into areas of Low, Medium and High catch based on data from commercial fishing logs (1995-1997) [11][29]. Stratified estimates were used until the assessment by Trzcinski et al.[26]. 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 [11][26]. Stratified random design reduces the variance in the mean of the estimate, therefore it was decided to transition to a stratified random survey design with more standardized protocols and expanded geographic coverage in 2016.

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 without 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. Catch Per Unit Effort (CPUE) ignores the effects of competition

for baited hooks within and between species and the problem of gear saturation [13]. Competition between target and non-target fish for the same hooks may result in gear saturation when the local population density of the target or non-target species exceeds the number of hooks which results in decreased efficiency of fishing effort [5]. It will also lead to bias in the abundance trends estimated using CPUE.

Target and non-target species will respond differently to temperature and depth and this may impact the catchability of longline gear. Over the scale of this large management unit (app 2000km long) temperature and depth will be associated with changes in fish assemblages thereby influencing hook competition. Temperature is also linked to the activity and feeding of ectotherms - which includes fish, amphibians and reptiles whose regulation of body temperature depends on external sources. In summary, there can be spatial and temporal variability in the abundance of target and non-target species and their catchability that could bias an index of abundance that ignores these sources of variability.

Stephen Smith [24] recommended using a multinomial model to account for the number of halibut caught, the number of hooks occupied by other species as well as empty-baited and empty-unbaited hooks and developed a package (`llsurv`) using the software R [21][23]. The `llsurv` package uses explicit formulae to compute maximum likelihood estimates of the multinomial model parameters in order to estimate relative abundance.

Here, we reproduce `llsurv` survey indices using a general purpose optimizer, so that the estimation method will be easier to modify later on. We then propose a novel model framework using Template Model Builder (TMB) [19] for model implementation, and `ggplot2` [28] for data visualization from within the R statistical environment [21]. In this new framework, we can include covariates such as depth and temperature. We also include a spatial random field in the model, evaluate the effectiveness of the stratification scheme, and test an alternative and more flexible method for reducing variance in estimates of abundance.

1.2 Data Description

The stratified random survey area is divided into 5 (4X5YZ, 4W, 4V, 3P, 3NO) area strata each with 3 (30-130 m, 131-250 m, 251-750 m) depth zones. Depth strata are selected based on exploratory analyses of catch rates by depth using fixed-stations and commercial index sets from the Atlantic Halibut survey. The depth stratum boundaries (30 - 750 m) were chosen because they contained most of the fishing sets and inferred Atlantic Halibut habitat. The depth stratification can also provide a proxy for temperature and some bottom habitat information [10]. For example, Smith [24] found that there is a strong correlation between depth and temperature for depths deeper than 280 m for sets in NAFO areas 4VWX. It is worthwhile to mention that the temperature for each longline set is reasonably constant during the fishing but depth can be variable for sets along a slope. In total there are 15 area-depth strata. Stations are randomly assigned to strata with the number of stations allocated proportionally by the size of each stratum. A total of 150 stations were allocated in 2017 and 153 stations were allocated in 2018, while in 2019 additional stations were added to strata with only 2 stations in an effort to reduce the probability of unfished strata or strata with only one station fished.

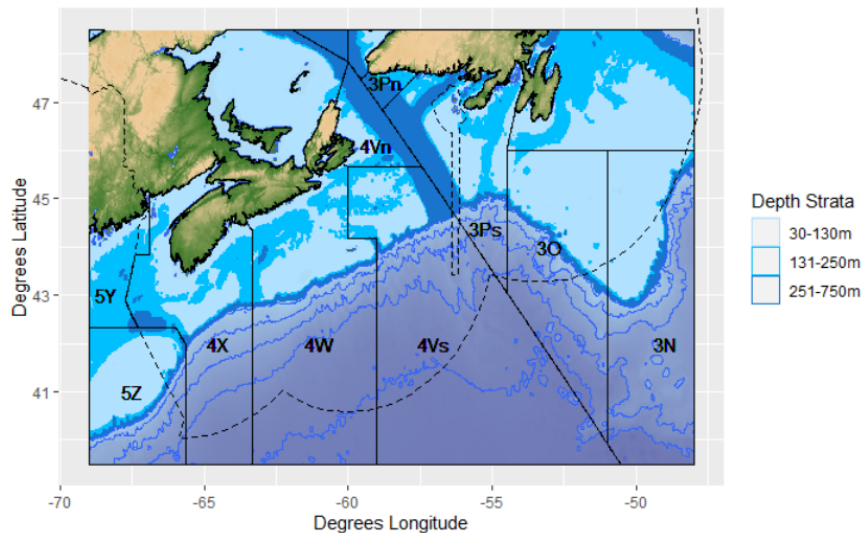


Figure 1.2: The survey area for the 3NOPs4VWX5Zc Atlantic Halibut stock (with five area strata: 4X5YZ, 4W, 4V, 3P and 3NO).

The Atlantic Halibut longline survey with a stratified random design generated the dataset considered here. Trained observers monitored the longline and recorded the conditions of a subset 300 hooks as they were retrieved. Specifically, each longline placed at the seafloor contained 1000 baited hooks and the hook conditions of 10 samples of 30 hooks were recorded for estimating hook occupancy. In addition, there exists a protocol to refish a station if the corresponding set is hauled back with no baited hooks. For each set we obtain counts of empty-unbaited hooks, empty-baited hooks, broken hooks, hooks with target catch and hooks with non-target catch. The corresponding strata information and soak time for each set were also included along with other essential information (Table 1.1 and Figures 1.3 to 1.5). For 2017 and 2018, model estimates of bottom temperature at time and location were obtained from http://marine.copernicus.eu/services-portfolio/access-to-products/?option=com_csw&view=details&product_id=GLOBAL_REANALYSIS_PHY_001_030 [14] by Dr. Zeliang Wang at BIO. The 2019 model estimates were not available for inclusion in this thesis. Issues that arose during the 2019 survey resulted in fewer stations being successfully completed (Table 1.2 and Figure 1.5).

STATION	STRATUM ID	SOAKMINP3P1	P1LONG
501	H11	584	-65.00200
502	H11	727	-64.71117
503	H11	628	-65.07950
504	H11	370	-64.78967
505	H11	360	-65.17850
P1LAT	P1DEPTH	EMPTY UNBAITED	EMPTY BAITED
45.45167	38	229	32
45.22783	46	245	15
45.14717	72	159	103
43.77800	32	37	257
43.34200	148	93	202
MISS BROKEN	HALIBUT	OTHER SPECIES	HOOKS SAMPLED
1	0	38	300
3	0	37	300
1	6	31	300
0	1	5	300
0	5	0	300
STRATUM ID ALT	MED TEMP	GLORYS TEMP	
1	11.37	7.18	
1	6.72	6.72	
1	10.13	NaN	
1	3.25	4.34	
1	6.24	6.24	

Table 1.1: Partial longline survey data for 2017. STATION indicates the survey station number. STRATUM ID and STRATUM ID ALT both indicate the corresponding stratum id (H11: area strata 4X5YZ and depth strata 30-130 m) for the survey station. SOAKMINP3P1, P1LONG and P1LAT represent the soak time and recorded latitude and longitude coordinate for the starting point of each longline set. Columns EMPTY UNBAITED, EMPTY BAITED, MISS BROKEN, HALIBUT and OTHER SPECIES are the count of hooks without bait, with bait, missing or broken, with halibut and other species at retrieval. HOOKS SAMPLED is the number of hooks sampled from the longline set. MED TEMP is the average winter bottom temperature and GLORYS TEMP contains temperature data obtained from Copernicus Marine Environment Monitoring Service (CMEMS) of the European Union (EU). NaN indicates a missing temperature.

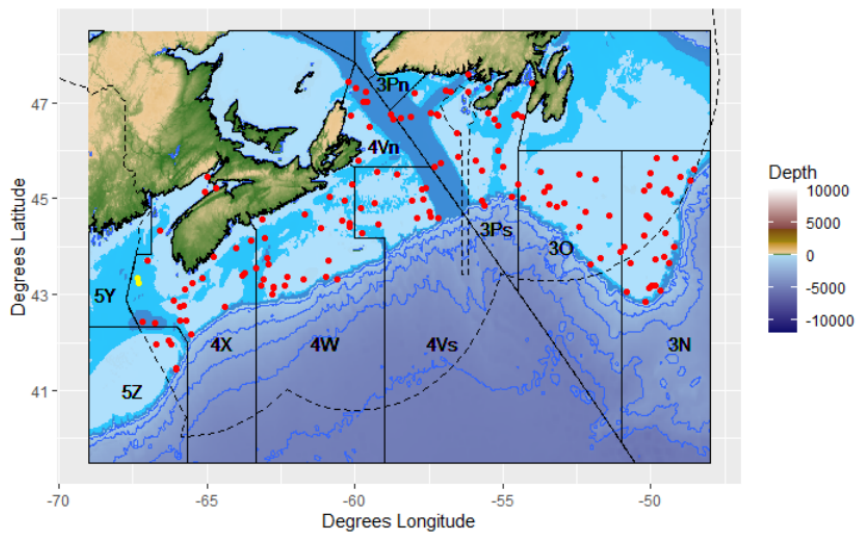


Figure 1.3: Locations of the 150 survey stations in 2017. Yellow points are stations 523 and 525 where only 30 hooks were sampled.

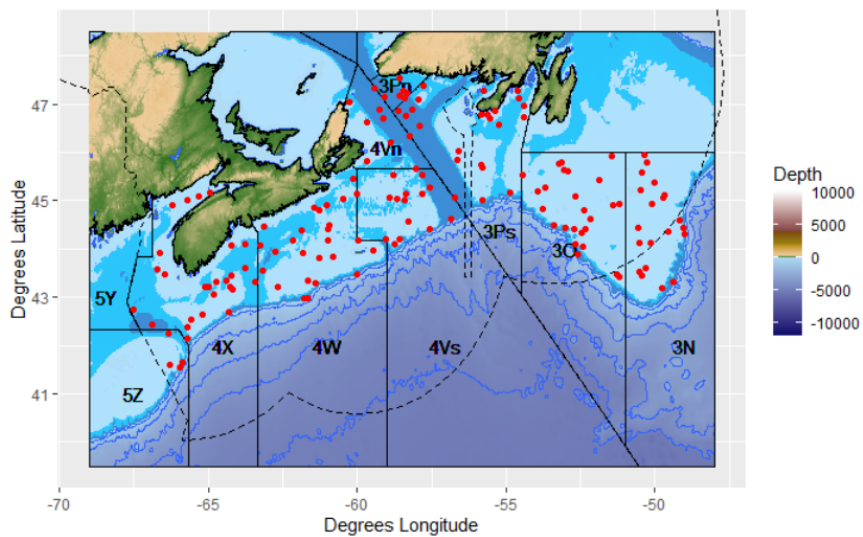


Figure 1.4: Locations of the 153 survey stations in 2018.

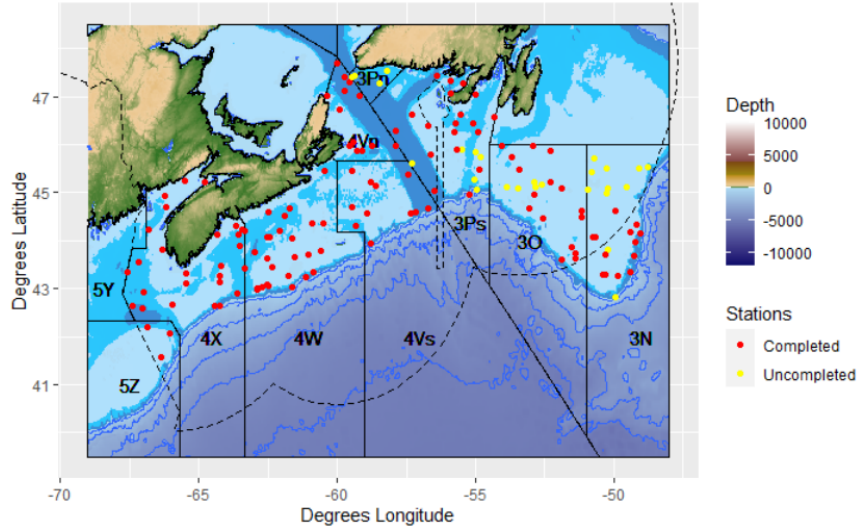


Figure 1.5: Locations of the 153 survey stations in 2019. Red points are the successfully completed stations and yellow points are the uncompleted stations due to issues that arose during the 2019 survey.

Stratum ID	Allocation of Survey Stations for 2017	Allocation of Survey Stations for 2018	Allocation of Survey Stations for 2019
H11	16 (16)	16 (16)	16 (16)
H12	11 (11)	11 (11)	11 (11)
H13	2 (2)	3 (3)	3 (3)
H21	15 (15)	15 (15)	15 (15)
H22	7 (7)	7 (7)	7 (7)
H23	2 (2)	3 (3)	3 (3)
H31	11 (11)	11 (11)	11 (11)
H32	5 (5)	5 (5)	5 (5)
H33	8 (8)	8 (8)	8 (8)
H41	13 (13)	13 (13)	10 (13)
H42	8 (8)	8 (8)	5 (8)
H43	8 (8)	8 (8)	4 (8)
H51	39 (39)	39 (39)	25 (39)
H52	2 (2)	3 (3)	2 (3)
H53	3 (3)	3 (3)	2 (3)
Total Number of Survey Stations	150 (150)	153 (153)	127 (153)

Table 1.2: The planned (in brackets) and completed allocation of survey stations per stratum and the total number of survey stations.

For stations 523 and 525 (with stratum id H12) in 2017 (Figure 1.3), the observer only recorded data for 30 hooks which was against observer protocol. Since there are 1000 hooks deployed per line and 10 samples of 30 hooks (300 hooks per station) need to be used to estimate hook occupancy, these two stations were data deficient. There are 11 stations in strata H12 (area 4X5YZ with depth 131-250 m) which suggests that removing these two incompletely observed stations will not have a great effect on subsequent analyses (Table 1.2). Therefore, stations 523 and 525 will be dropped from the 2017 analyses.

The missing depths for stations 570, 572 and 573 in 2018 and stations 570, 571, 572, 573 and 574 in 2019 were replaced by the depth estimated using publicly available bathymetric data from the United States National Atmospheric and Oceanographic Administration (NOAA). Depths were extracted from the bathymetric layer using the function `getNOAA.bathy` from the R package `marmap` [20] and the coordinate bounds of the survey area were used to import bathymetric data from the NOAA server. Then the depth of the area of interest was obtained from the imported bathymetric data using the `get.depth` function from the R package `marmap`. The bathymetric method is used because it is accepted as the best available measurement when depth is not directly measured or recorded during setting of the longline by the onboard observer. 62% of the absolute differences between survey depth and estimated depth are within 10 meters (Figure 1.6 through Figure 1.8). Stations with large depth variation ($> |10|$ m) are in areas where there are substantial depth slopes and these rapid depth changes are not well described by the bathymetric models (Figure 1.9 through Figure 1.11).

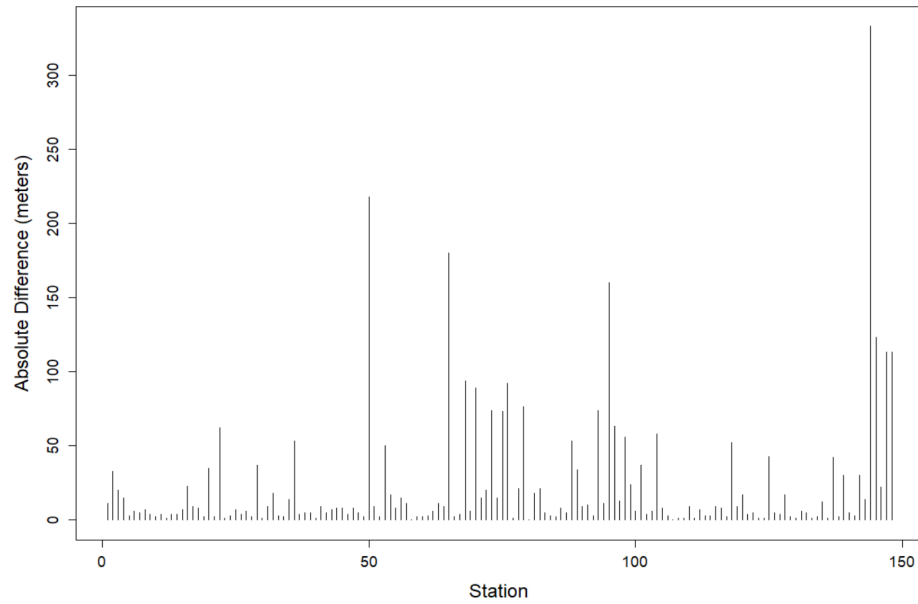


Figure 1.6: Absolute differences (m) between survey depth and depth obtained using the bathymetric method for 2017.

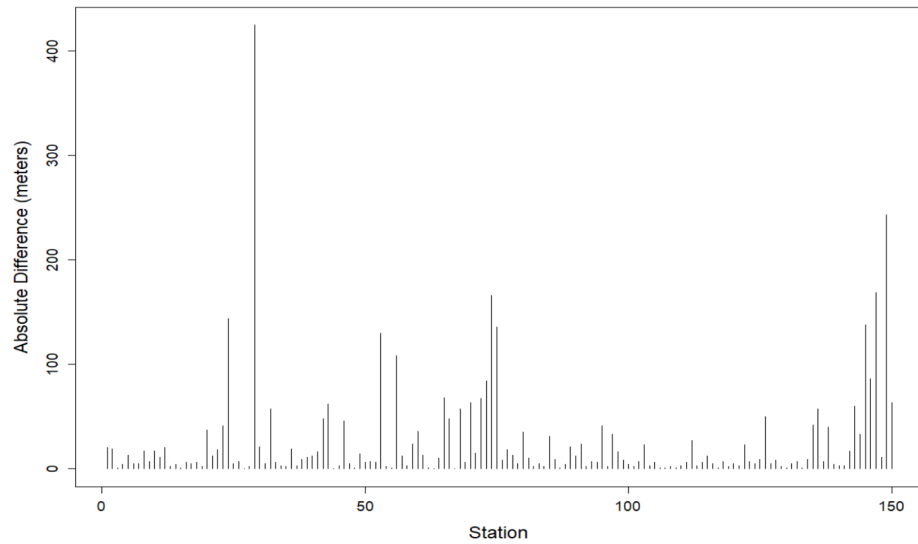


Figure 1.7: Absolute differences (m) between survey depth and depth obtained using the bathymetric method for 2018 without stations 570, 572 and 573.

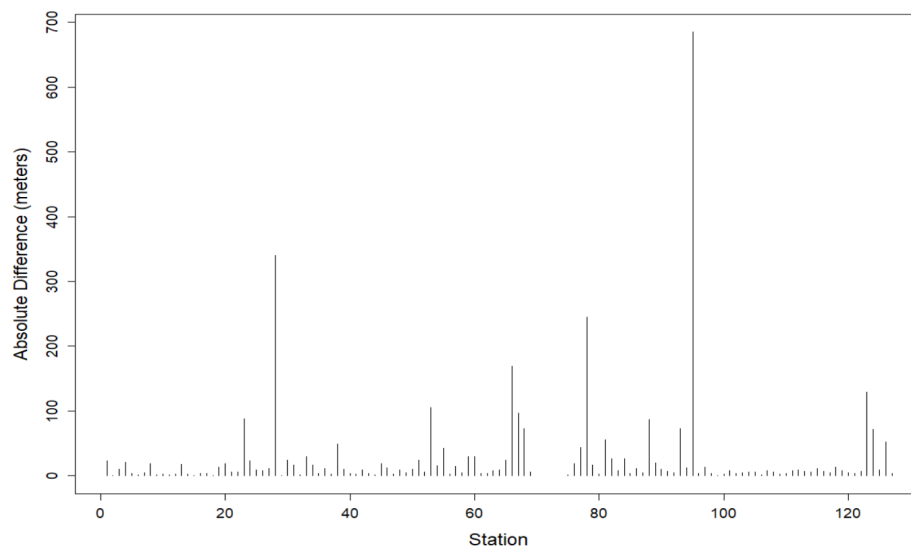


Figure 1.8: Absolute differences (m) between survey depth and depth obtained using the bathymetric method for 2019 without stations 570, 571, 572, 573 and 574.

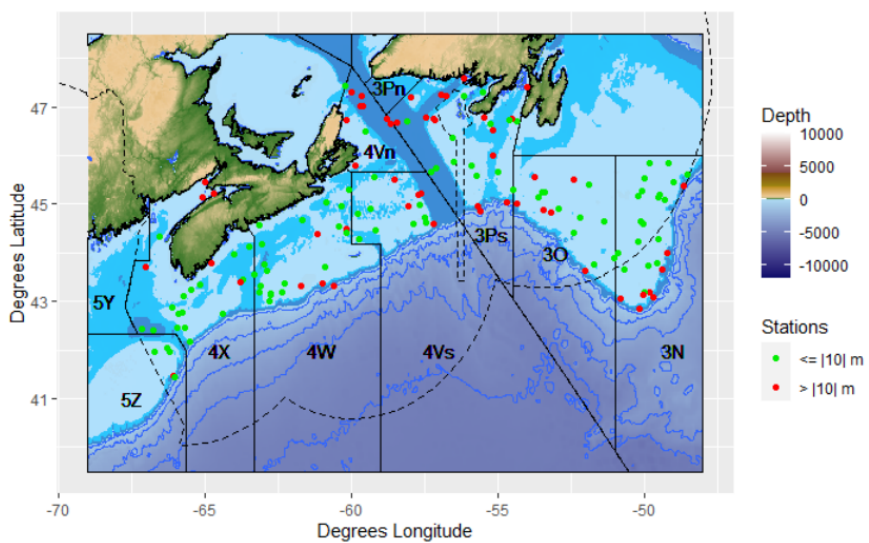


Figure 1.9: Locations of the survey stations with indication of absolute difference between survey depth and bathymetric depth for 2017. Red points are stations with absolute difference greater than 10 m.

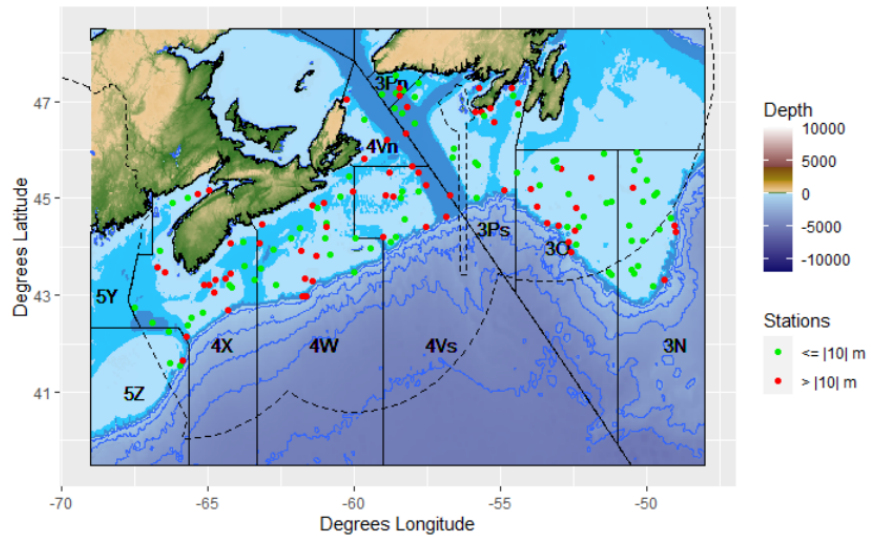


Figure 1.10: Locations of the survey stations with indication of absolute difference between survey depth and bathymetric depth for 2018. Red points are stations with absolute difference greater than 10 m.

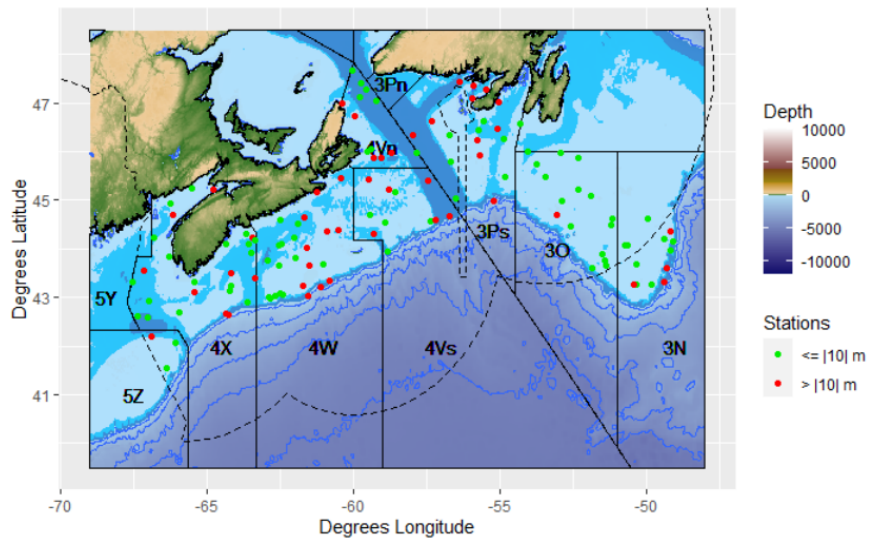


Figure 1.11: Locations of the survey stations with indication of absolute difference between survey depth and bathymetric depth for 2019. Red points are stations with absolute difference greater than 10 m.

From 2017 to 2019, 70% to 90% of the sets in the longline survey have temperature data derived from attached VEMCO minilogs (Table 1.3). The model generated bottom temperature data (GLORYS12V1) at time and location were used for survey stations which did not have temperature data from the minilogs. The GLORYS12V1 from Copernicus Marine Environment Monitoring Service (CMEMS) of the European Union (EU) is an eddy-resolving global ocean product, which assimilates observational data (e.g. satellite sea surface temperature data, satellite altimeter data, temperature and salinity data from ARGO profilers and ship surveys). The GLORYS12V1 reanalysis product has a horizontal resolution of $1/12^\circ$, and 50 vertical levels [14]. The model estimated bottom temperature is used because the observed temperature and generated temperature are well correlated (Figure 1.12). Though the model tends to overestimate or underestimate the temperatures near Shelf edges, the residuals between the observed and estimated temperatures are small (Figures 1.13 to 1.15).

Year	Number of Stations with Temperature	Number of Stations without Temperature	Percentage of Stations with Temperature
2017	134	14	90.54%
2018	125	28	81.70%
2019	101	16	79.53%

Table 1.3: Summary of the number of stations with and without observed temperature data and the percentage of stations with observed temperature.

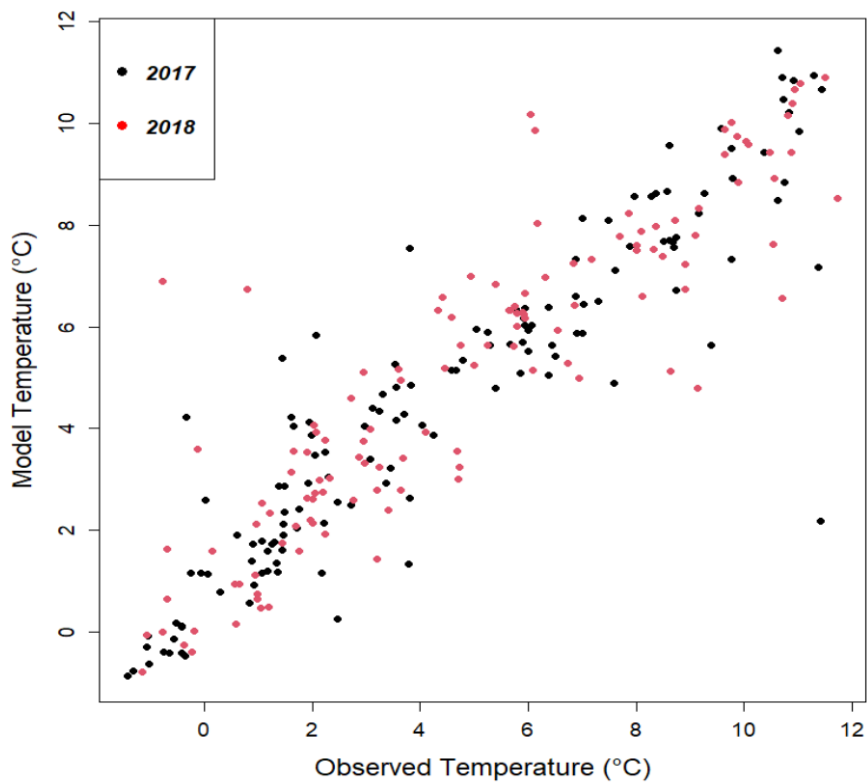


Figure 1.12: Observed temperature and model estimated temperature for 2017 and 2018.

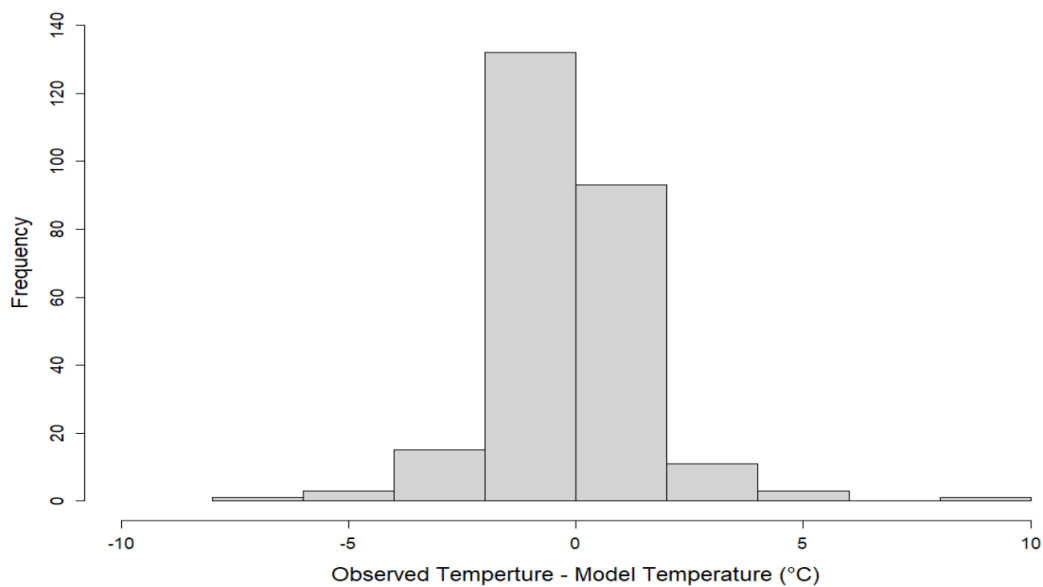


Figure 1.13: The residuals between observed temperature and model estimated temperature for 2017 and 2018.

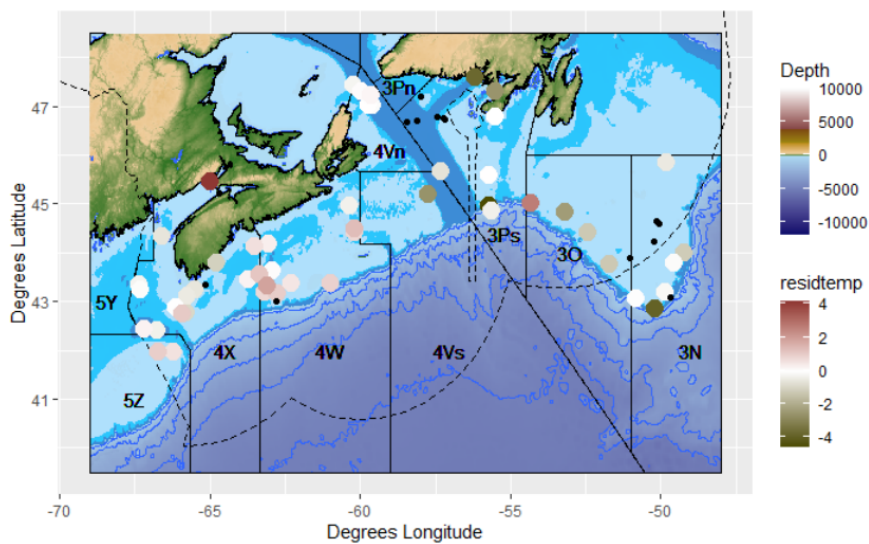


Figure 1.14: The residuals between observed temperature and model estimated temperature for 2017. The small black points are the stations with missing temperature.

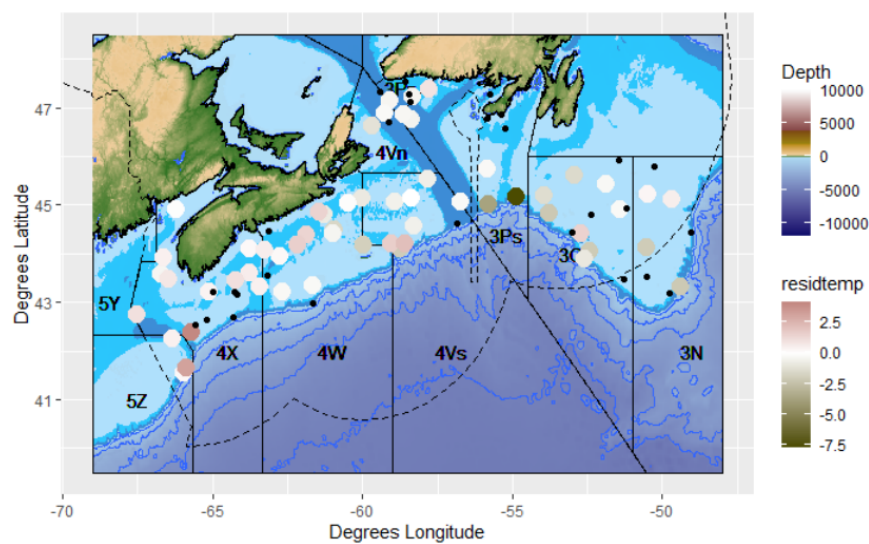


Figure 1.15: The residuals between observed temperature and model estimated temperature for 2018. The small black points are the stations with missing temperature.

It can be noticed from Table 1.1 that soak times in the survey have varied over survey stations. However from Figure 1.16 we can observe that the distribution of soak time does not vary over years. Comparing to the other depth strata, there are more longline sets with soak time below 450 min in depth strata 30-130 m (Figure 1.17). The distribution of soak time in different area strata varies but there is no obvious pattern for soak time in different area strata (Figure 1.18).

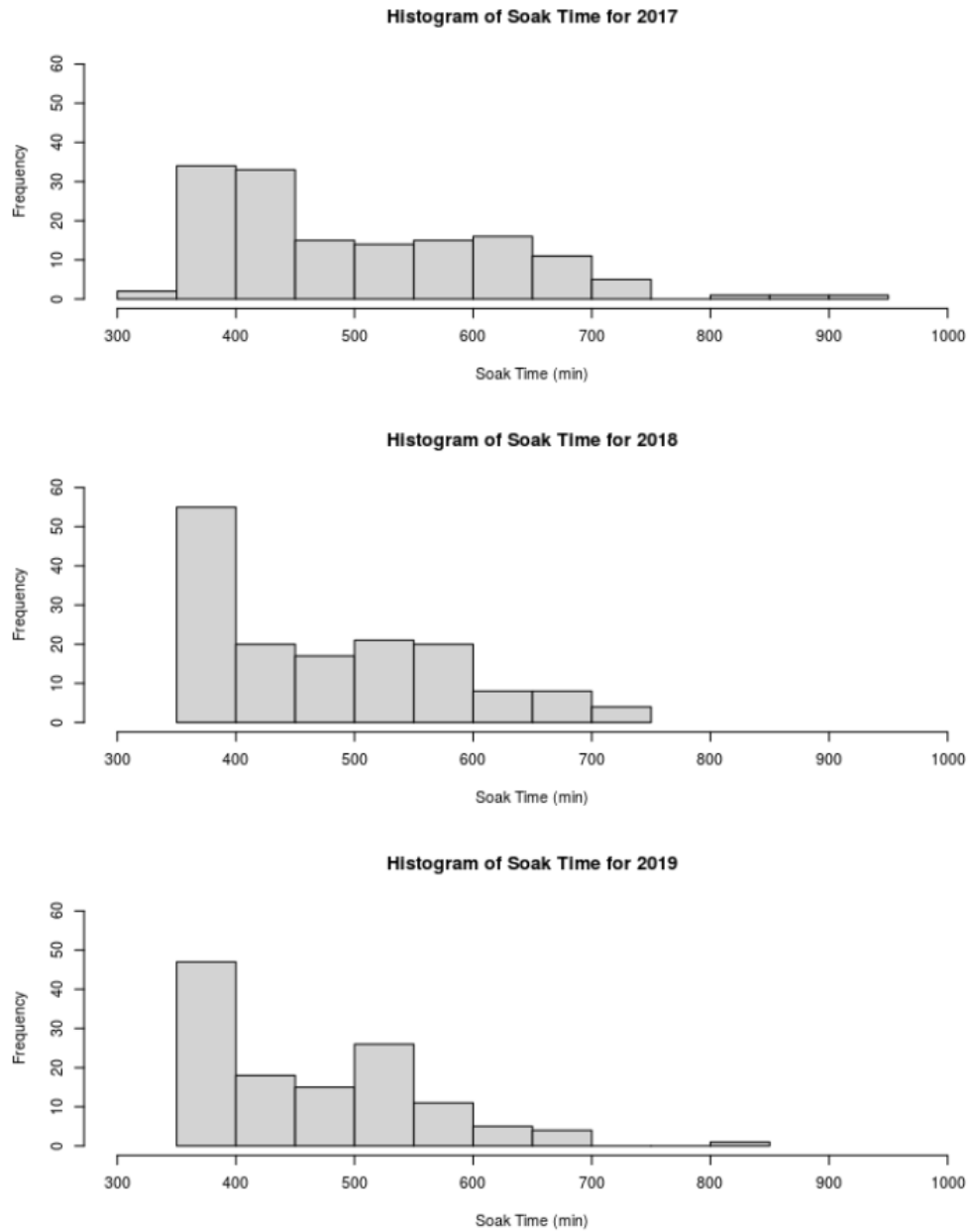


Figure 1.16: Histograms of the soak time by year for 2017 to 2019.

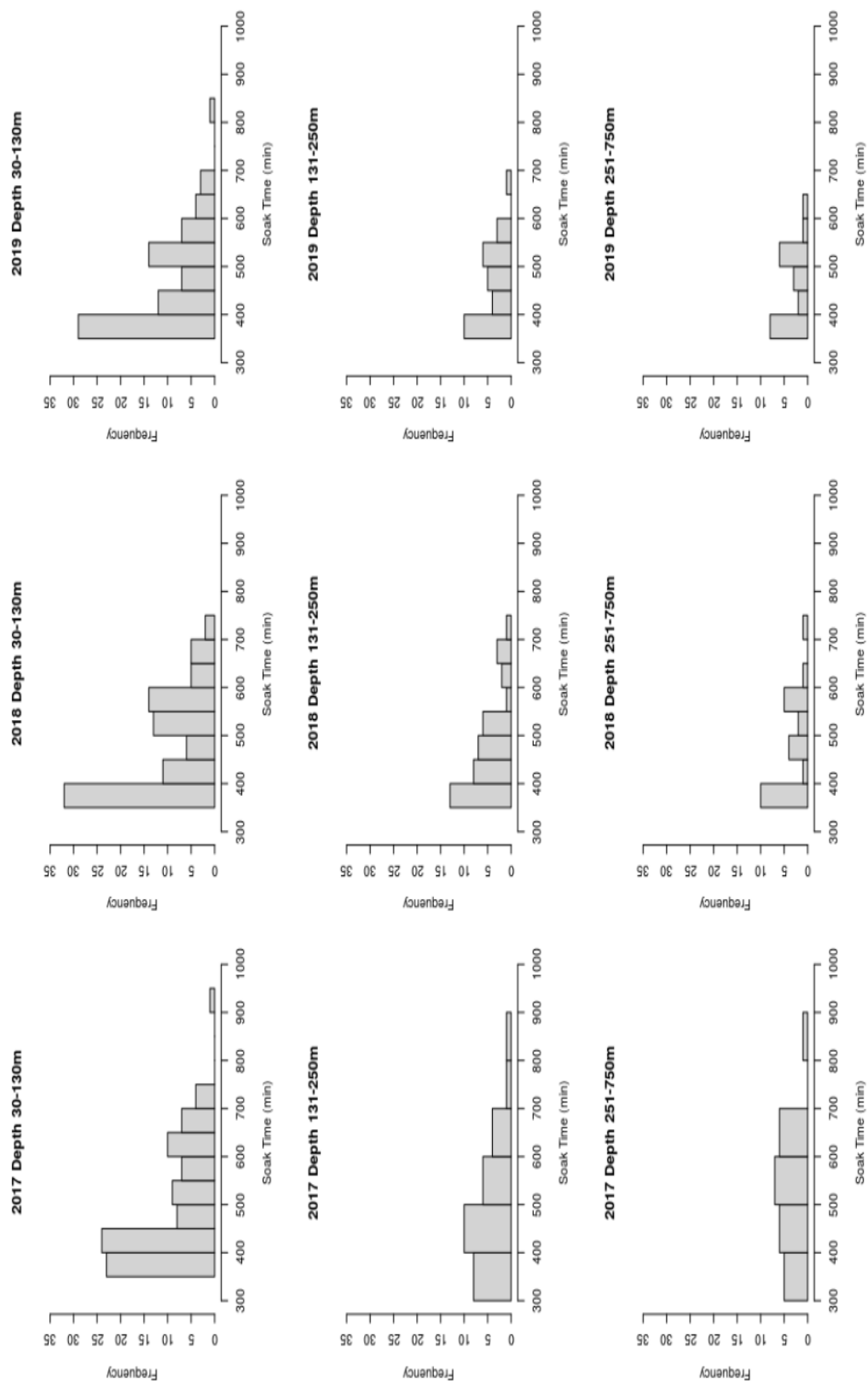


Figure 1.17: Histograms of the soak time by depth strata for 2017 to 2019.

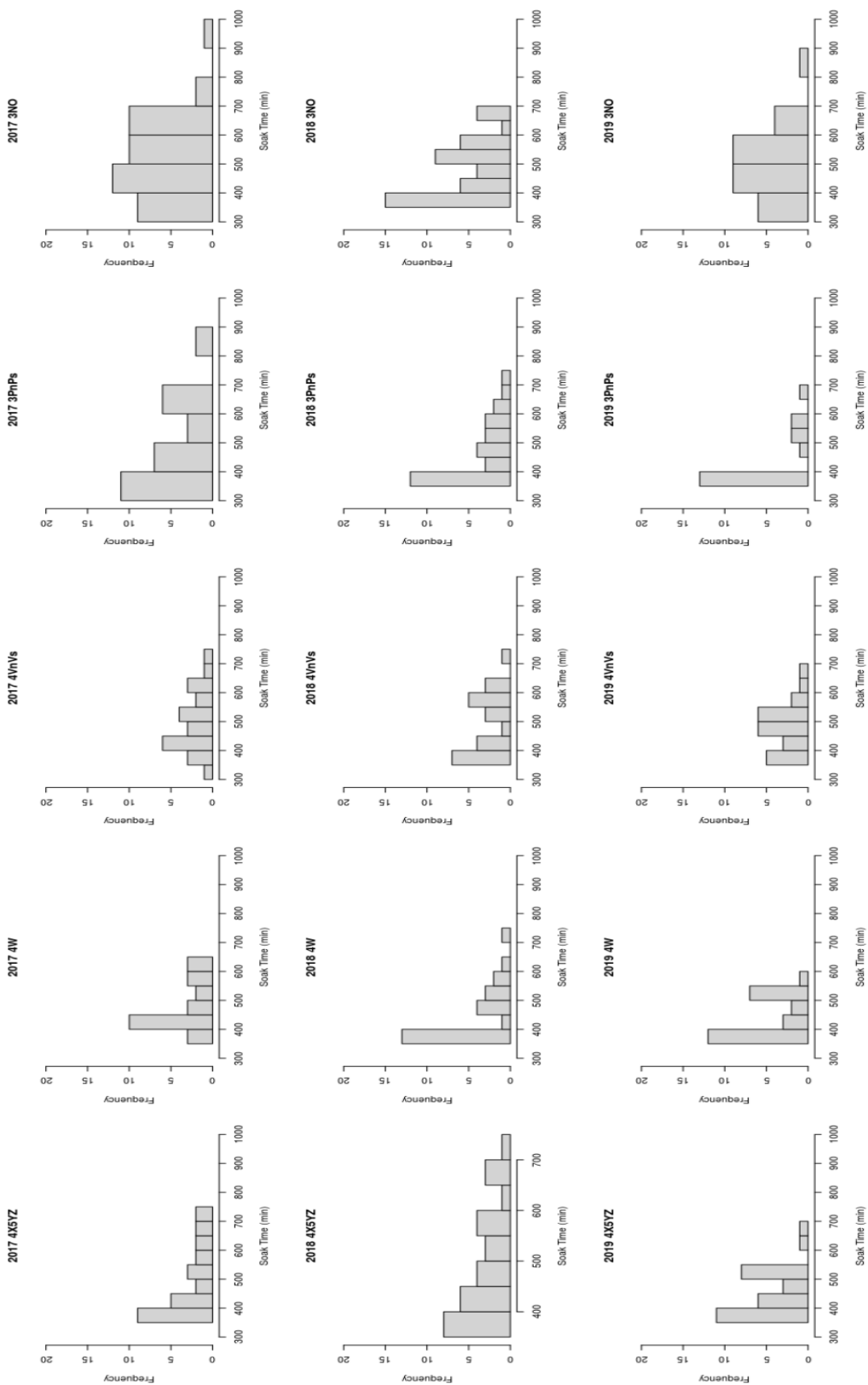


Figure 1.18: Histograms of the soak time by area for 2017 to 2019.

1.3 Survey Index Reproduction

1.3.1 Multinomial Exponential Model

The multinomial exponential model (MEM) originally proposed by Rothschild [22] and modified by Etienne et al. [13] is currently used by DFO for the Atlantic Halibut longline survey.

Let T_T be the time of catching a target species (halibut) on a hook and assume T_T follows an exponential distribution with rate λ_T . Then the probability of catching a halibut before time u is

$$P(T_T \leq u) = 1 - e^{-\lambda_T u}.$$

Let T_{NT} be the time it takes for a non-target species to be caught and assume T_{NT} follows an exponential distribution with rate λ_{NT} . The probability distribution of the time it takes for either a target or non-target species to be caught, $T = \min\{T_T, T_{NT}\}$ is also exponential with rate $\lambda = \lambda_T + \lambda_{NT}$.

The possible states for a hook after soak time S upon retrieval onboard are,

- $\{I = 0\} = \{\text{The hook is still baited}\}$ with probability

$$P(I = 0) = P(T > S) = e^{-\lambda S},$$

- $\{I = Target\} = \{\text{The hook has a target species}\}$ with probability

$$P(I = Target) = P(T_T < T_{NT} | T < S)P(T < S) = \frac{\lambda_T}{\lambda}(1 - e^{-\lambda S}),$$

- $\{I = NTarget\} = \{\text{The hook has a non-target species}\}$ with probability

$$P(I = NTarget) = P(T_{NT} < T_T | T < S)P(T < S) = \frac{\lambda_{NT}}{\lambda}(1 - e^{-\lambda S}),$$

where I is an indicator variable.

Assuming that all the hooks on a longline are independent, the likelihood for hooks being baited or not baited and if not baited containing a halibut or another

species is given by

$$L(\lambda_T, \lambda_{NT}) = \binom{N}{N_B} \binom{N_T + N_{NT}}{N_T} (e^{-\lambda S})^{N_B} \\ \times (1 - e^{-\lambda S})^{N_T + N_{NT}} \left(\frac{\lambda_T}{\lambda}\right)^{N_T} \left(\frac{\lambda_{NT}}{\lambda}\right)^{N_{NT}},$$

where

- N is the number of hooks on the longline,
- N_B is the number of baited hooks at the end of the soak time,
- N_T is the number of the target species caught,
- N_{NT} is the number of the non-target species caught.

This model is called the Multinomial Exponential Model (MEM) [22]. The expected numbers of the halibut caught by a longline with N hooks is N times the probability $P(I = Target)$ for a halibut being caught when there is competition from other species. Therefore the number of the halibut that were observed reflects an index of the abundance of that species modified by the relative abundance of other species competing for the same hooks. Parameters λ_T and λ_{NT} reflect the relative abundance of target and non-target species respectively.

It is common for some hooks to return with no bait and no fish at the end of soak time in longline fishing. The MEM was modified by Etienne et al. [13] to account for these empty-unbaited hooks so that more precise estimates of annual indices of relative abundance could be obtained. It's assumed that all empty-unbaited hooks are caused by the escape of fish. That is,

- The number N_B of baited hooks out of N total hooks at the end of the soak time is binomially distributed with probability of success $e^{-\lambda S}$:

$$N_B \sim \mathcal{B}(N, e^{-\lambda S}).$$

- Given the number of empty-unbaited hooks ($N - N_B$), the number of halibut caught is $N_T + N_E^{(T)}$ and also has a binomial distribution:

$$N_T + N_E^{(T)} | N - N_B \sim \mathcal{B}\left(N - N_B, \frac{\lambda_1}{\lambda}\right),$$

where $N_E^{(T)}$ is the number of caught halibut escaping.

- Given $N_T + N_E^{(T)}$, the total number of halibut on the longline is N_T and is binomially distributed:

$$N_T | N_T + N_E^{(T)} \sim \mathcal{B} \left(N_T + N_E^{(T)}, (1 - p_T) \right),$$

where p_T is the escapement rate of caught target fish.

- Given $N_{NT} + N_E^{(NT)}$, the total number of non-target species on the longline is N_{NT} and also follows a binomial distribution:

$$N_{NT} | N_{NT} + N_E^{(NT)} \sim \mathcal{B} \left(N_{NT} + N_E^{(NT)}, (1 - p_{NT}) \right),$$

where $N_E^{(NT)}$ is the number of caught non-target species escaping and p_{NT} is the escaping rate of caught non-target fish.

$N_E^{(NT)}$ and $N_E^{(T)}$ are missing quantities but the sum N_E of these two quantities is observed. Therefore the likelihood function of the model is:

$$\begin{aligned} L(\lambda_T, \lambda_{NT}, p_T, p_{NT}) &= \frac{N!}{N_B! N_T! N_{NT}! N_E!} (e^{-\lambda_S})^{N_B} (1 - e^{-\lambda_S})^{N - N_B} \\ &\times \left(\frac{\lambda_T}{\lambda} (1 - p_T) \right)^{N_T} \left(\frac{\lambda_{NT}}{\lambda} (1 - p_{NT}) \right)^{N_{NT}} \left(\frac{\lambda_T p_T + \lambda_{NT} p_{NT}}{\lambda} \right)^{N_E}. \end{aligned}$$

This is a multinomial distribution with parameters N and α :

$$(N_B, N_T, N_{NT}, N_E) \sim \mathcal{M}(N, \alpha)$$

where N is the number of hooks on a longline set and $\alpha = (e^{-\lambda_S}, (1 - e^{-\lambda_S}) \frac{\lambda_T}{\lambda} (1 - p_T), (1 - e^{-\lambda_S}) \frac{\lambda_{NT}}{\lambda} (1 - p_{NT}), (1 - e^{-\lambda_S}) \frac{\lambda_T p_T + \lambda_{NT} p_{NT}}{\lambda})$.

There were two models proposed by Etienne et al. [13] based on two different situations. The first model (MEM1) assumed that empty-unbaited hooks were caused by non-target species taking the bait and escaping with the corollary that the probability of escape for target species would be 0 ($p_T = 0$). The second model (MEM2) assumed that empty-unbaited hooks came from either target or non-target species. In other words, the escaping rate is the same for target and non-target catch ($p_T = p_{NT}$). However, those authors who have investigated the impact of bait loss on longline

estimates have generally assumed that missing baits were caused by non-target fish [7][17][27]. Since the longline survey is designed for catching and retaining the target species in terms of bait, hook type and size, depth, and timing, non-capture would be more likely for other species in the area. In addition, the target species would also have lower abundance than all other possibly caught species. Moreover, assigning the empty-unbaited hooks to species other than halibut will at worst lead to an underestimation of the halibut count [13]. Therefore, MEM1 was selected and the likelihood function was re-written as follows:

$$L(\lambda_T, \lambda_{NT}, p_{NT}) = \frac{N!}{N_B! N_T! N_{NT}! N_E!} (e^{-\lambda S})^{N_B} (1 - e^{-\lambda S})^{N - N_B} \\ \times \left(\frac{\lambda_T}{\lambda}\right)^{N_T} \left(\frac{\lambda_{NT}}{\lambda} (1 - p_{NT})\right)^{N_{NT}} \left(\frac{\lambda_{NT} p_{NT}}{\lambda}\right)^{N_E}.$$

The maximum likelihood estimators of MEM1 with a constant soak time are given by:

$$\hat{\lambda}_T = \frac{N_{T+}}{N_+ - N_{B+}} \frac{1}{S} \log\left(\frac{N_+}{N_{B+}}\right) \quad (1.1)$$

$$\hat{\lambda}_{NT} = \frac{N_{NT+} + N_{E+}}{N_+ - N_{B+}} \frac{1}{S} \log\left(\frac{N_+}{N_{B+}}\right) \quad (1.2)$$

$$\hat{p}_{NT} = \frac{N_{E+}}{N_{E+} + N_{NT+}} \quad (1.3)$$

where $N_{B+} = \sum_{l=1}^L N_{B_l}$, $N_{T+} = \sum_{l=1}^L N_{T_l}$, $N_{NT+} = \sum_{l=1}^L N_{NT_l}$, $N_{E+} = \sum_{l=1}^L N_{E_l}$ and l is the number of longline sets. Additionally, empty-unbaited hooks also include broken and missing hooks.

1.3.2 Template Model Builder (TMB)

We coded the MEM1 proposed by Etienne et al. [13] in TMB by first constructing the negative log-likelihood function in R. After ensuring that the negative log-likelihood function in R was working appropriately, the function was converted into C++ code and the TMB package was used for estimation so that the MEM1 could be efficiently implemented using simple C++ templates. The model was formulated in C++ and

the data were manipulated in R. The function `MakeADFun` from the R package `TMB` was used to construct objective functions and pass the objective function value to the optimizer in R for the final estimation. This package is used because it combines external libraries `CppAD`, `Eigen` and `CHOLMOD` to obtain an efficient implementation of the applied Laplace approximation with exact derivatives and can speed up a process significantly when the number of iterations is high. For more details about the `TMB` package please refer to the `TMB` reference manual by Kristensen et al. [18]. In addition, R codes used for analysis and data visualization are provided in the Appendices.

1.3.3 Comparison of Results

The R package `llsurv` [24] uses explicit formulae for maximum likelihood estimates. These only allow one value for soak time therefore the mean of the soak time for the survey was used to reproduce results. The identical results obtained using two different methods and packages (Table 1.4) indicate that survey indices estimated using the explicit method can be successfully reproduced using a general purpose optimizer. For the year 2017, the estimated relative abundance indices with soak time fixed at 500.5068 minutes for the target ($\hat{\lambda}_T$) and non-target species ($\hat{\lambda}_{NT}$) are 1.843592e-05 and 0.001904881 respectively. The estimated survey indices with constant soak time (478.1895 minutes) for the target ($\hat{\lambda}_T$) and non-target species ($\hat{\lambda}_{NT}$) are 2.661856e-05 and 0.002068632 respectively in 2018. The probability of escape for the non-target species (\hat{p}_{NT}) is 0.8908173. For 2019, the estimated relative abundance indices with fixed soak time (463.2283 minutes) for the target ($\hat{\lambda}_T$) and non-target species ($\hat{\lambda}_{NT}$) are 2.324046e-05 and 0.002227046 respectively.

The `TMB` generated indices of abundance can also be estimated with set-specific soak time (Table 1.5). The survey indices estimated using the set-specific soak time will be more accurate as the Multinomial model assumes that catch increases linearly with soak time. However error rates between the results obtained using fixed soak time and set-specific time are small. The error rate is obtained using the difference

between the survey indices estimated with fixed soak time and set-specific soak time and divided by the survey index estimated with set-specific soak time. Specifically, error rates between the estimated survey indices for target species with mean of soak time and set-specific soak time are approximately 1.90% in 2017, 0.37% in 2018 and 0.23% in 2019. A comparison of the proportion of hooks with halibut from the survey with soak time indicates that soak time does not appear to have any effect on the proportion of halibut caught. A similar conclusion can be drawn from the proportion of hooks with catches of the other non-target species as well (Figure 1.19).

Year	Mean of the Soak Time (mins)	$\hat{\lambda}_T$ (llsurv)	$\hat{\lambda}_{NT}$ (llsurv)	$\hat{\lambda}_T$ (TMB)	$\hat{\lambda}_{NT}$ (TMB)
2017	500.5068	1.843592e-05	0.001904881	1.843592e-05	0.001904881
2018	478.1895	2.661856e-05	0.002068632	2.661856e-05	0.002068632
2019	463.2283	2.324046e-05	0.002227046	2.324046e-05	0.002227046

Table 1.4: The estimated relative abundance indices across the entire survey area with mean of the soak time for the survey for target and non-target species using R packages TMB and llsurv for hook occupancy data from the halibut longline survey for the years 2017, 2018 and 2019.

Year	$\hat{\lambda}_T$	$\hat{\lambda}_{NT}$	\hat{p}_{NT}
2017	1.878635e-05	0.001941087	0.8795759
2018	2.671808e-05	0.002076371	0.8908173
2019	2.329181e-05	0.002232086	0.8963697

Table 1.5: The estimated relative abundance indices across the entire survey area with set-specific soak time for target and non-target species and the escaping probability of caught non-target species using the TMB package for hook occupancy data from the halibut longline survey for the years 2017, 2018 and 2019.

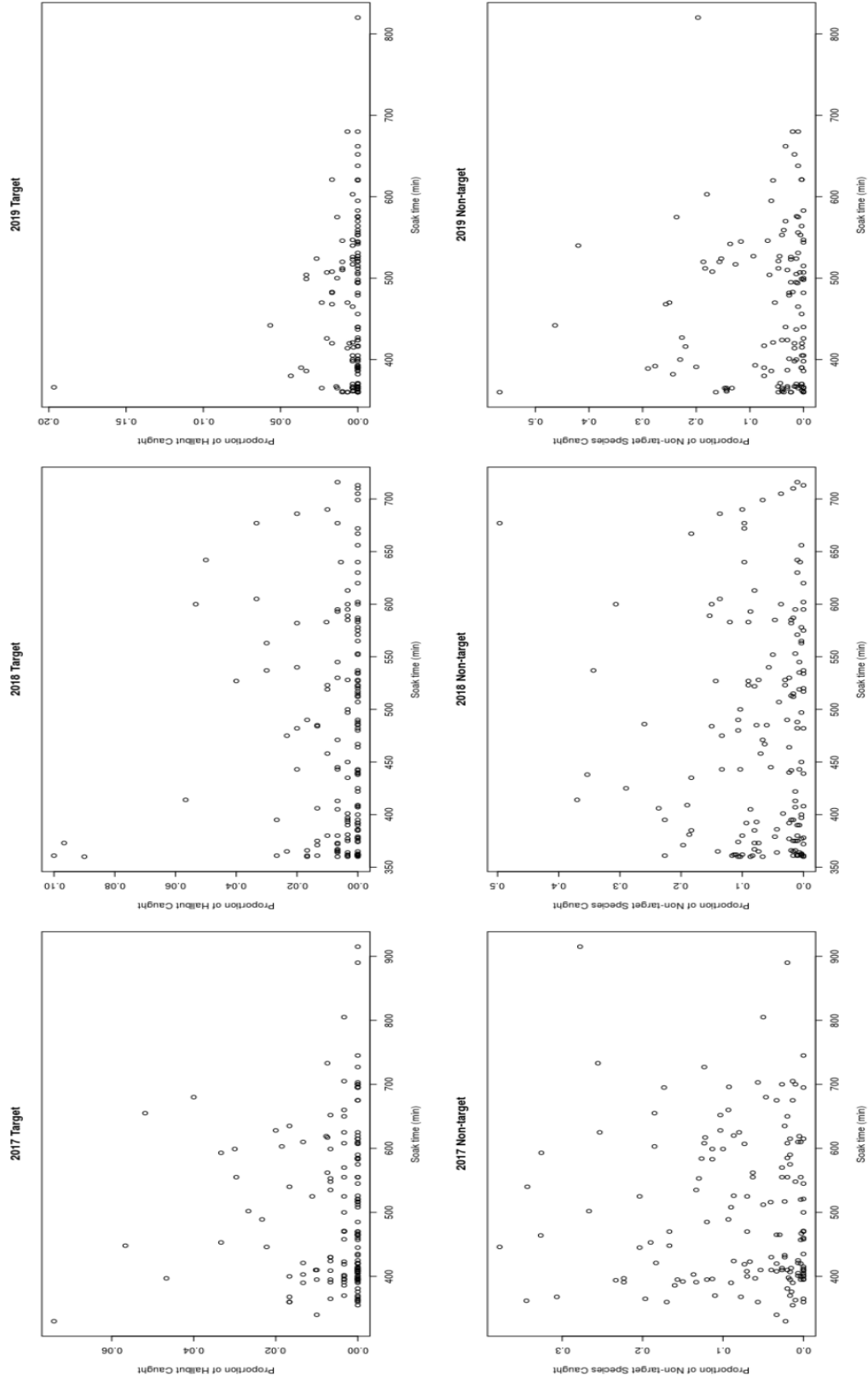


Figure 1.19: A comparison of the proportion of halibut and non-target species caught with different soak times for 2017, 2018 and 2019.

Chapter 2

Spatial Models and Variable Selection

Here we start with a discussion of the spatial patterns found in the survey data. The following two sections then focus on the application of spatial statistics to the current survey indices estimation model and culminate with the proposal of a novel model formulation. Variable selection for this model is conducted in the final section.

2.1 Spatial Patterns

Typically when a region exhibits spatial structure, the points closer to each other are more strongly correlated than the distant points. The maps of the proportion of baited and unbaited hooks in each set show obvious spatial patterns (Figures 2.1 to 2.3). Here unbaited hooks include missing and broken hooks as well as hooks that came back with halibut, other species or without bait. There are more unbaited hooks in the depth zone 251-750 m and more baited hooks in the depth zone 30-130 m. With regard to the area strata, there are more unbaited hooks in areas 4X5XY, 4W and 3P. This result is further supported by ANOVA tests of the proportion of unbaited hooks by strata (Tables 2.1 to 2.3). The p-values ($9.4e-08$ and $9.81e-07$) for these tests are very small indicating that there are significant differences in the proportion of unbaited hooks by strata. To attempt to deal with this spatial structure a spatial random field is considered. A spatial random field is a random function over an arbitrary domain meaning that it is a function $f(x)$ that takes on a random value at each point $x \in \mathbb{R}^n$. It is also defined as a generalization of a stochastic process taking values in a Euclidean space, and defined over a parameter space of at least one dimension [3]. Abundance estimates obtained with inclusion of a spatial random field will be compared to those previously obtained based on the stratification scheme.

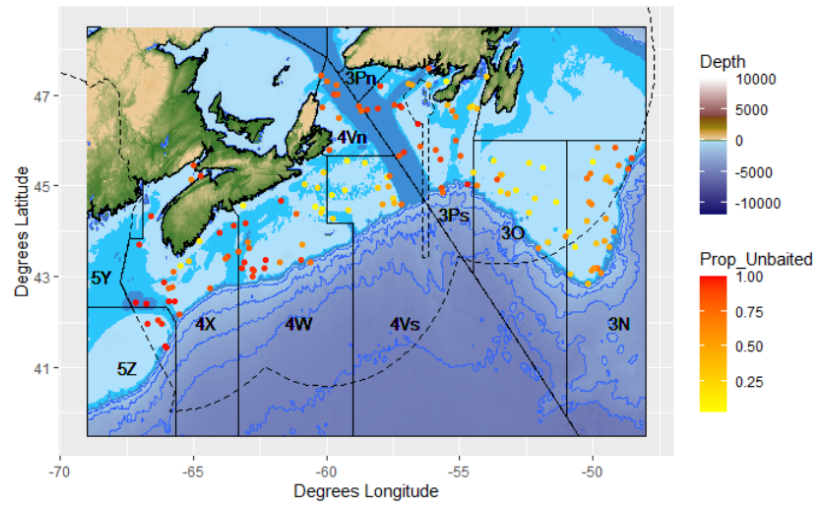


Figure 2.1: The distribution of unbailed hooks in 2017. Here unbailed hooks include missing and broken hooks as well as hooks that came back with halibut, other species or without bait.

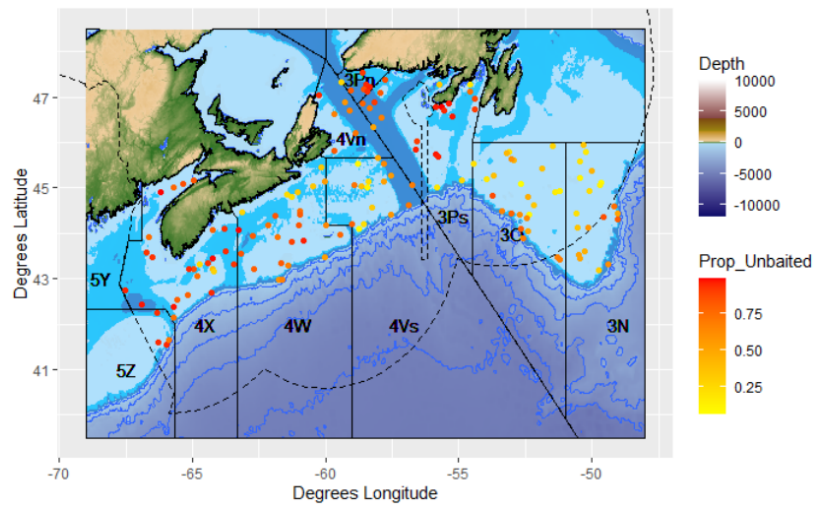


Figure 2.2: The distribution of unbailed hooks in 2018. See caption for Figure 2.1.

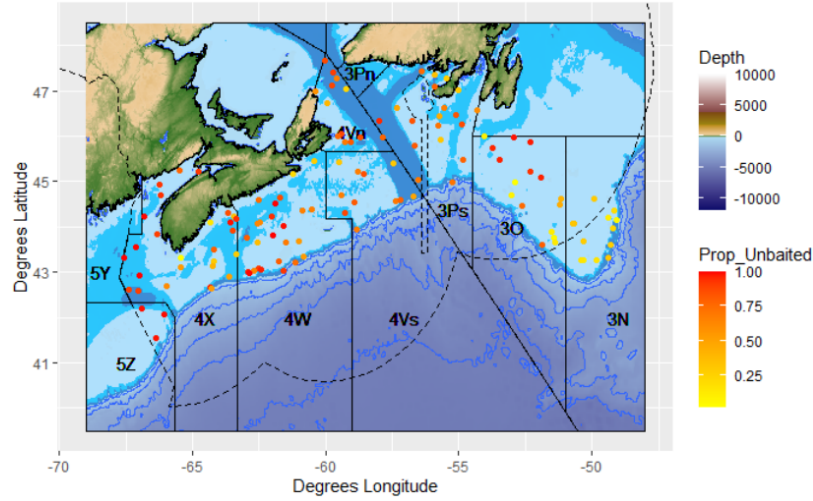


Figure 2.3: The distribution of unbaited hooks in 2019. See caption for Figure 2.1.

	DF	sum Sq	Mean Sq	F value	Pr(>F)
STRATUM ID	14	4.452	0.3180	4.323	2.77e-06
Residuals	133	9.783	0.0736		

Table 2.1: Summary of the ANOVA test of proportion of unbaited hooks by strata for 2017.

	DF	sum Sq	Mean Sq	F value	Pr(>F)
STRATUM ID	14	4.850	0.3465	6.693	2.98e-10
Residuals	138	7.143	0.0518		

Table 2.2: Summary of the ANOVA test of proportion of unbaited hooks by strata for 2018.

	DF	sum Sq	Mean Sq	F value	Pr(>F)
STRATUM ID	14	2.986	0.21328	2.92	0.000834
Residuals	112	8.180	0.07304		

Table 2.3: Summary of the ANOVA test of proportion of unbaited hooks by strata for 2019.

2.2 Models with a Gaussian Random Field

2.2.1 Binomial Model with Gaussian Random Field (glmmTMB)

The number of unbaited hooks ($N_{UNB,i}$), which includes missing and broken hooks as well as hooks that came back with halibut, other species and without bait, on a longline set at the end of the soak time is binomially distributed

$$N_{UNB,i} \sim \mathcal{B}(N_i, \pi_i).$$

The logit of the probability π_i of obtaining unbaited hooks on a longline set can be modelled as a linear function of soak time (x_{i_soak}), depth (x_{i_depth}) and temperature (x_{i_temp}) at each survey location. That is

$$\text{logit}(\pi_i) = \beta_0 + \beta_1 \times x_{soak_i} + \beta_2 \times x_{depth_i} + \beta_3 \times x_{temp_i}, \quad (2.1)$$

and the binomial model with Gaussian random field is defined as

$$\text{logit}(\pi_i) = \beta_0 + \beta_1 \times x_{soak_i} + \beta_2 \times x_{depth_i} + \beta_3 \times x_{temp_i} + \omega_i \quad (2.2)$$

where ω_i is a Gaussian random field with Matérn covariance function $C(h)$. A Matérn covariance is generally used to define the statistical covariance between measurements made at two locations that are h units distant from each other. The covariance is stationary as it only depends on distances between locations. It is also isotropic when distance is Euclidean distance. A Matérn covariance function is used because it is flexible and includes both the exponential and Gaussian covariance functions as special cases, moving between them via a smoothness parameter. The Matérn covariance function [16] is defined as

$$C(h) = \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}} \left(\frac{h}{\phi}\right)^\nu K_\nu\left(\frac{h}{\phi}\right),$$

where

- K_ν : modified Bessel function of the second kind,
- σ^2 : variance/sill parameter,

- $\phi > 0$: range parameter \Rightarrow effective/practical range (the distance at which the correlation decreases to 0.05),
- $\nu > 0$: smoothness parameter \Rightarrow mean square differentiability of the process (Hausdorff/fractal dimension of Gaussian sample paths).
- h : standardized distance matrix of the longline survey location,

Special cases:

$$\nu = 0.5 \Rightarrow \text{Exponential covariance function: } C(h) = \sigma^2 \exp\left(-\frac{h}{\phi}\right)$$

$$\nu = \infty \Rightarrow \text{Gaussian covariance function: } C(h) = \sigma^2 \exp\left(-\frac{h^2}{\phi^2}\right)$$

The estimation of π_i can be performed using package Generalized Linear Mixed Models using Template Model Builder (glmmTMB) [9] as it fits linear and generalized linear mixed models with TMB and has built-in spatial covariance options. glmmTMB fits models using maximum likelihood estimation via TMB and random effects are integrated out using the Laplace approximation.

The probability of getting unbaited hooks was predicted at 1000 randomly selected locations from the block id which is the unique identifier for each grid cell of the entire survey area. The grid is made up of 2 km \times 2 km blocks, with the point indicated being the center of the block. The same 1000 points were used for all three predictions. The predictions were obtained using mean of the soak time for the survey and without taking temperature into account as the soak time and temperature data for the block id are not easily available.

Similar to spatial patterns displayed by observed data (Figures 2.1 through 2.3), the probability of getting unbaited hooks is higher in areas 4X5YZ, 4W and 3P (Figures 2.4 through 2.6). Though glmmTMB has built-in spatial covariance options, the family object in function glmmTMB does not contain the multinomial model which allows for estimating the relative abundance indices for halibut directly. The binomial model can only estimate the probability of getting unbaited hooks but the goal is to estimate the survey indices for halibut. Therefore instead of using glmmTMB, the multinomial exponential model (MEM1) with a Gaussian random field will be coded in C++ and used for analysis.

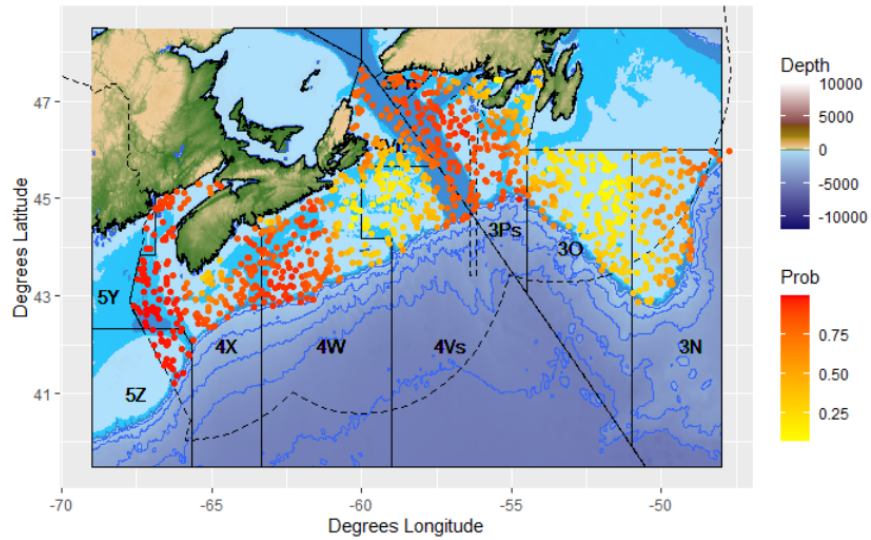


Figure 2.4: The predicted probabilities of getting unbaited hooks at 1000 randomly selected locations for 2017.

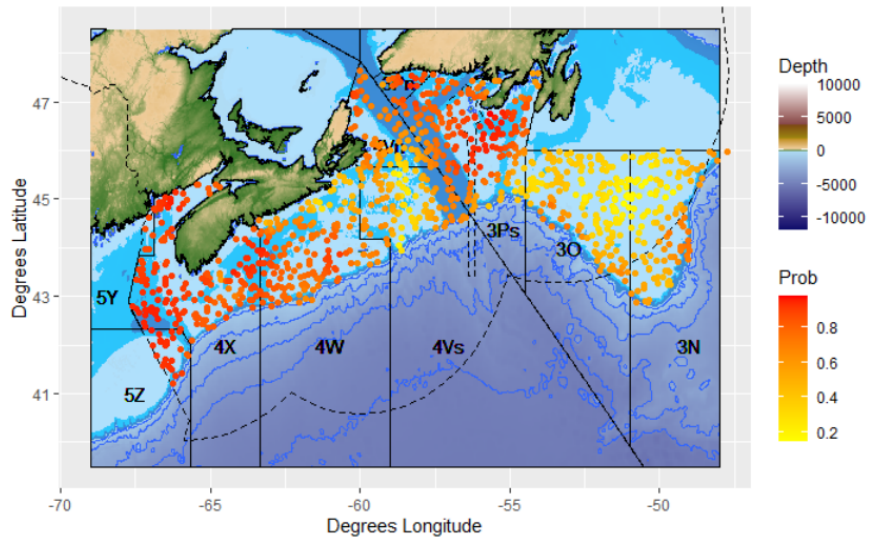


Figure 2.5: The predicted probabilities of getting unbaited hooks at 1000 randomly selected locations for 2018.

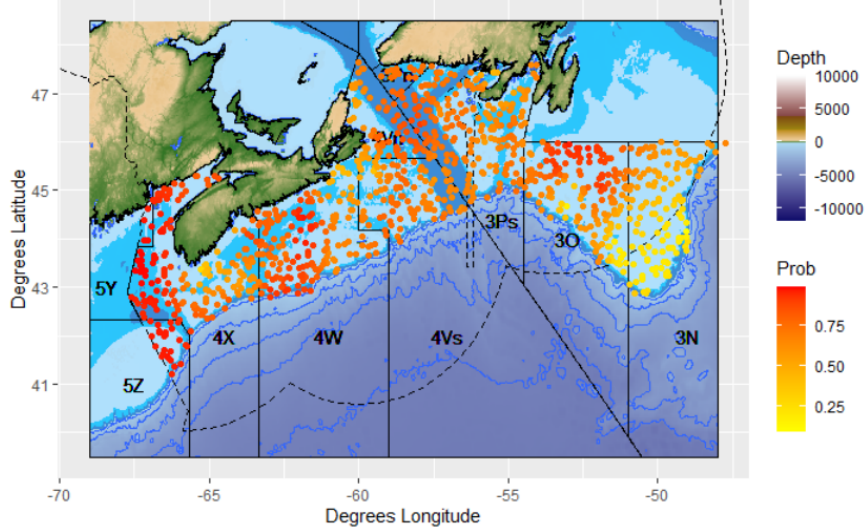


Figure 2.6: The predicted probabilities of getting unbaited hooks at 1000 randomly selected locations for 2019.

2.2.2 Multinomial Exponential Model (MEM1) with Gaussian Random Field (TMB)

The relative abundance index λ_i at each survey station is affected by the Gaussian random field ω_i with Matérn covariance function, therefore we will modify the MEM1 described in the previous section by adding a random field to both target and non-target species. The smoothness parameter ν of the Matérn covariance function is typically fixed at 1 since it is difficult to estimate. The initial value of the range parameter ϕ is the area of the standardized coordinates of the survey stations. The relative abundance indices for target and non-target species at each survey station are defined as

$$\hat{\lambda}_{T.i} = \exp(\beta_{T.0} + \omega_{T.i}) \quad (2.3)$$

$$\hat{\lambda}_{NT.i} = \exp(\beta_{NT.0} + \omega_{NT.i}) \quad (2.4)$$

There are two scenarios for $\omega_{T.i}$ and $\omega_{NT.i}$, the first one is that they are from the same random field (Figure 2.7). This means that $\omega_{T.i}$ and $\omega_{NT.i}$ share one covariance function, so one choice of parameters ϕ , σ and ν evaluated for two different sets of ω . The other one is that $\omega_{T.i}$ and $\omega_{NT.i}$ are from different random fields meaning that

ω_{T_i} and ω_{NT_i} each has a covariance function, so two sets of parameters ϕ , σ and ν evaluated for two different sets of ω (Figures 2.8 and 2.9). By comparing the plots of correlation matrices for target and non-target species under the second situation for 2017 (Figures 2.8 and 2.9), it can be observed that patterns of correlation matrices for target and non-target species are quite different. Therefore, the model with ω_{T_i} and ω_{NT_i} from different random fields is used.

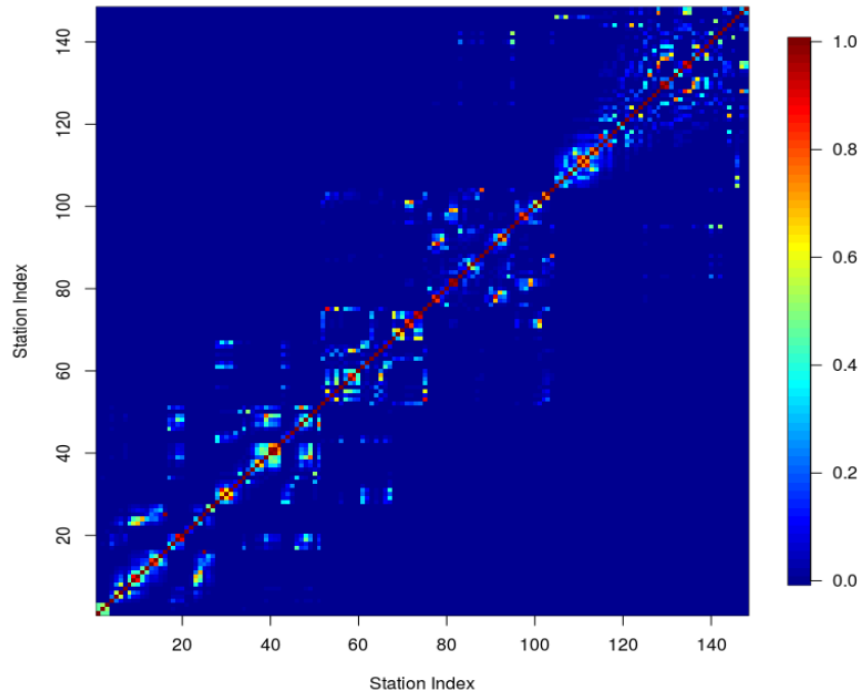


Figure 2.7: The correlation matrix for both target and non-target species for 2017. ω_{T_i} and ω_{NT_i} are from same random field.

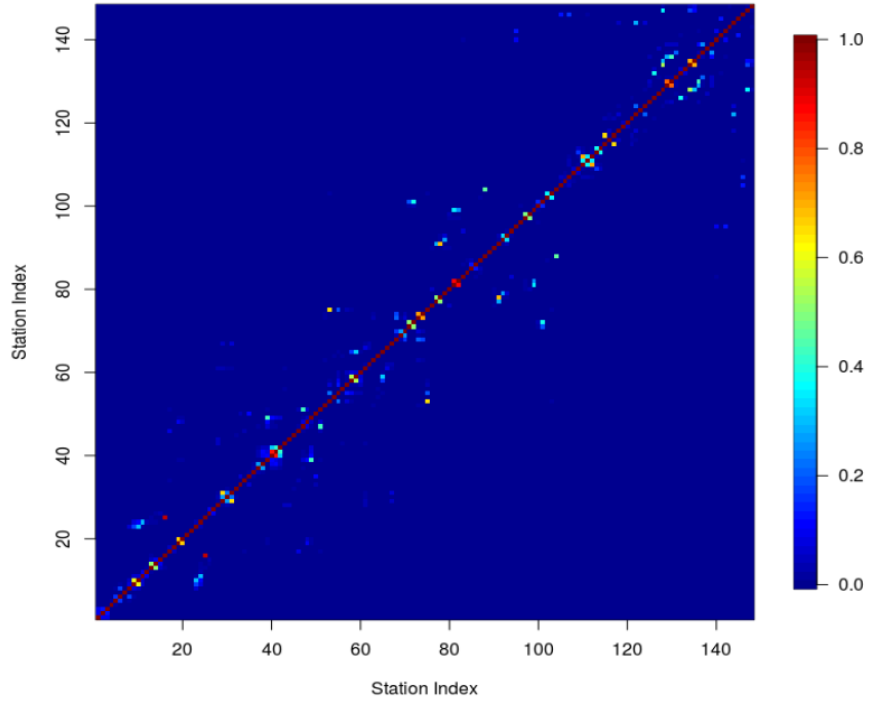


Figure 2.8: The correlation matrix for target species for 2017. $\omega_{T,i}$ and $\omega_{NT,i}$ are from different random fields.

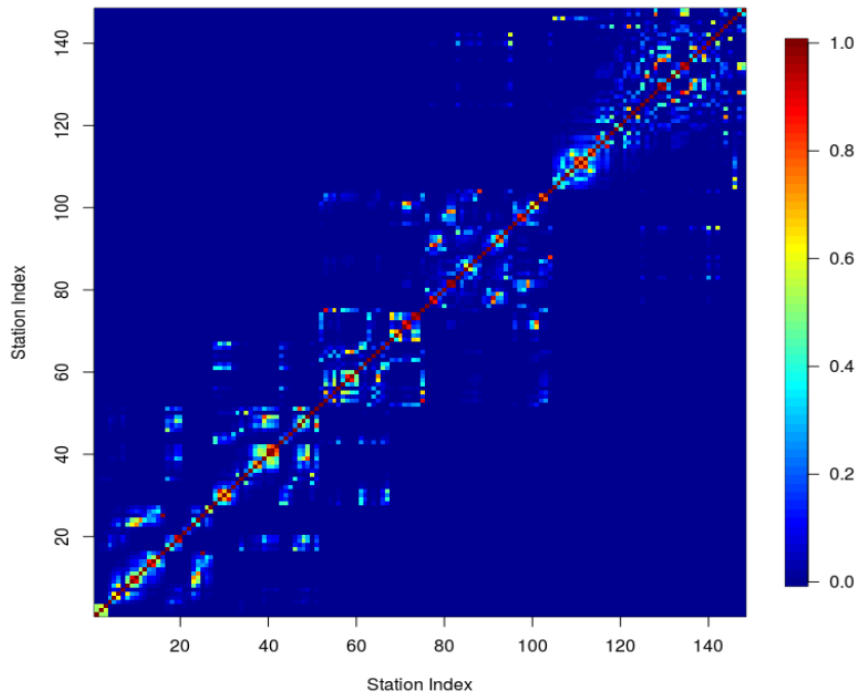


Figure 2.9: The correlation matrix for non-target species for 2017. $\omega_{T,i}$ and $\omega_{NT,i}$ are from different random fields.

2.2.3 Variable Selection

Covariates for depth, temperature and area strata (along with their potential interactions) are also introduced into the model (Equations 2.5 and 2.6). Since these may affect the relative abundance indices of target and non-target species

$$\hat{\lambda}_{T.i} = \exp(\beta_{T.0} + \omega_{T.i} + \beta_{T.1} \times x_{1.i} + \beta_{T.2} \times x_{2.i} + \beta_{T.3} \times x_{1.i} \times x_{2.i} + \dots) \quad (2.5)$$

$$\hat{\lambda}_{NT.i} = \exp(\beta_{NT.0} + \omega_{NT.i} + \beta_{NT.1} \times x_{1.i} + \beta_{NT.2} \times x_{2.i} + \beta_{NT.3} \times x_{1.i} \times x_{2.i} + \dots). \quad (2.6)$$

From the simplest model (the model without covariates) to the most complex model (the model with temperature, area strata, depth strata and interactions between area and depth strata), there are 14 models to be compared (Tables 2.4 through 2.6). Both Bayesian information criterion (BIC) and Akaike information criterion (AIC) are used and the model is selected based on BIC scores where

$$BIC = k \ln(n) - 2 \ln(L(\hat{\theta}))$$

and

$$AIC = 2k - 2 \ln(L(\hat{\theta})),$$

where

- n is the sample size,
- k is the number of parameters in the model plus the intercept,
- θ is the set of all parameters,
- $L(\hat{\theta})$ represents the maximized likelihood of the model under consideration.

AIC and BIC are criteria for model selection, the optimal model is selected based on the minimum AIC or BIC. BIC is used as the criteria for model selection because AIC penalizes the number of parameters less strongly than the BIC meaning that AIC tends to choose model with more parameters which may cause overfitting.

Model	Minimized Negative Log-likelihood	K	AIC	BIC
With Temperature	3692.523	4	7393.046	7405.035
With Depth + Temp	3692.381	6	7396.762	7414.745
With Temp + Depth Strata (3)	3692.088	8	7400.177	7424.155
Without Covariates	3712.183	2	7428.366	7434.361
With Area Strata (5) + Depth + Temp	3683.438	14	7394.875	7436.836
With Depth	3709.150	4	7426.300	7438.289
With Depth Strata (3)	3710.374	4	7428.748	7440.737
With Area Strata (5) + Depth	3692.886	12	7409.771	7445.738
With Area Strata (5) + Depth Strata (3) + Temp	3683.370	16	7398.740	7446.696
With Area Strata (5) + Depth Strata (3)	3693.860	14	7415.719	7457.680
With Area Strata (5) * Depth Strata (3) + Temp	3667.764	32	7399.527	7495.438
With Area Strata (5) * Depth	3704.438	20	7448.877	7508.821
With Area Strata (5) * Depth Strata (3)	3679.827	30	7419.653	7509.570
With Area Strata (5) * Depth + Temp	3786.990	22	7617.980	7683.919

Table 2.4: AIC and BIC for different models for 2017. The models are ranked according to their BIC scores in increasing order. Parameter for random field is not included when calculating AIC and BIC. K contains the parameters for both target and non-target species. There are 5 area strata and 3 depth strata. The symbol “*” represents the interactions between the covariates.

Model	Minimized Negative Log-likelihood	K	AIC	BIC
With Temperature	3597.365	4	7202.729	7214.851
With Depth + Temp	3597.191	6	7206.382	7224.565
With Temp + Depth Strata (3)	3595.191	8	7206.383	7230.626
With Area Strata (5) + Depth + Temp	3585.981	14	7199.961	7242.387
With Area Strata (5) + Depth Strata (3) + Temp	3584.606	16	7201.211	7249.698
With Depth Strata (3)	3616.249	4	7240.497	7252.619
Without Covariates	3621.410	2	7246.820	7252.881
With Depth	3619.843	4	7247.685	7259.807
With Area Strata (5) + Depth	3604.551	12	7233.103	7269.468
With Area Strata (5) + Depth Strata (3)	3600.357	14	7228.714	7271.140
With Area Strata (5) * Depth Strata (3) + Temp	3575.792	32	7215.584	7312.558
With Area Strata (5) * Depth Strata (3)	3594.379	30	7248.758	7339.671
With Area Strata (5) * Depth	3637.866	20	7315.733	7376.341
With Area Strata (5) * Depth + Temp	3656.634	22	7357.267	7423.937

Table 2.5: AIC and BIC for different models for 2018. See caption for Table 2.4.

Model	Minimized Negative Log-likelihood	K	AIC	BIC
Without Covariates	3447.512	2	6899.023	6904.712
With Depth Strata (3)	3447.203	4	6900.565	6911.942
With Depth	3446.283	4	6902.407	6913.784
With Area Strata (5) + Depth	3437.861	12	6899.723	6933.853
With Area Strata (5) + Depth Strata (3)	3436.680	14	6901.360	6941.178
With Area Strata (5) * Depth	3439.862	20	6919.724	6976.607
With Area Strata (5) * Depth Strata (3)	3422.570	30	6905.139	6990.465

Table 2.6: AIC and BIC for different models for 2019. See caption for Table 2.4.

We can observe from Tables 2.4 through 2.6 that the model to be selected is the model with only temperature which is defined as

$$\hat{\lambda}_{T.i} = \exp(\beta_{T.0} + \beta_{T.1} \times x_{temp.i} + \omega_{T.i}) \quad (2.7)$$

$$\hat{\lambda}_{NT.i} = \exp(\beta_{NT.0} + \beta_{NT.1} \times x_{temp.i} + \omega_{NT.i}). \quad (2.8)$$

However, temperature will not be included in the model used to provide an index of abundance because it is not possible to estimate the temperature data for sets that did not have temperature collected in 2019 at this time (Table 1.3). Additionally, the differences between the survey indices for halibut estimated by models with and without temperature are small. This indicates that while temperature is statistically significant it does not have a great impact on the predictions (Figures 2.10 and 2.11). By comparing the BIC scores of the models in Tables 2.4 to 2.6 the model with the lowest BIC and without temperature is the model without covariate. The model is then reduced to

$$\hat{\lambda}_{T.i} = \exp(\beta_{T.0} + \omega_{T.i}) \quad (2.9)$$

$$\hat{\lambda}_{NT.i} = \exp(\beta_{NT.0} + \omega_{NT.i}). \quad (2.10)$$

It also can be observed from Tables 2.4 to 2.6 that strata are not statistically important in the spatial model.

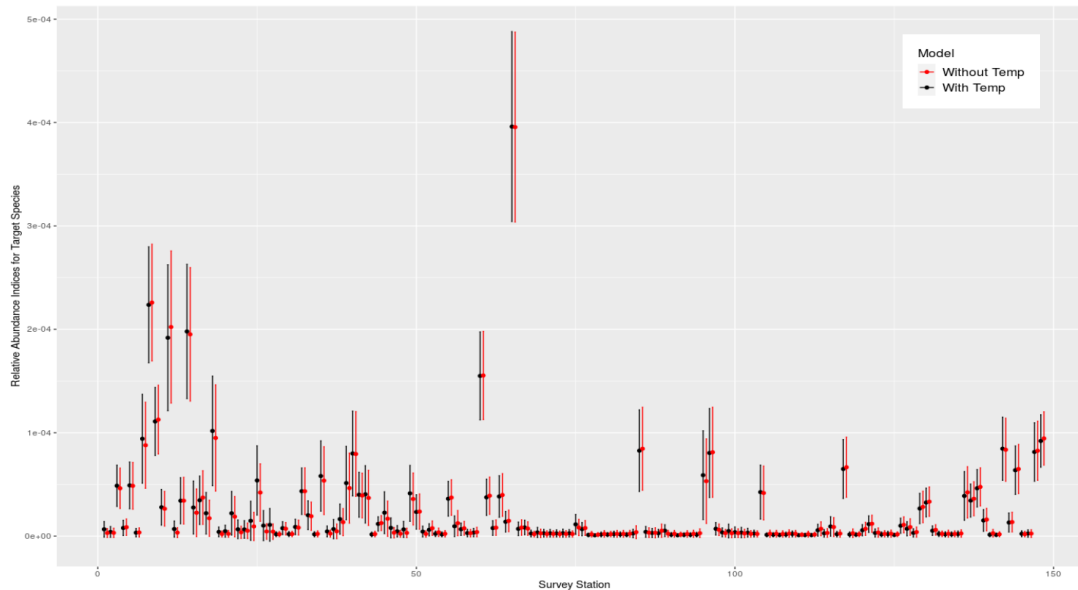


Figure 2.10: Relative abundance indices for halibut and corresponding standard errors estimated by models with and without temperature for 2017.

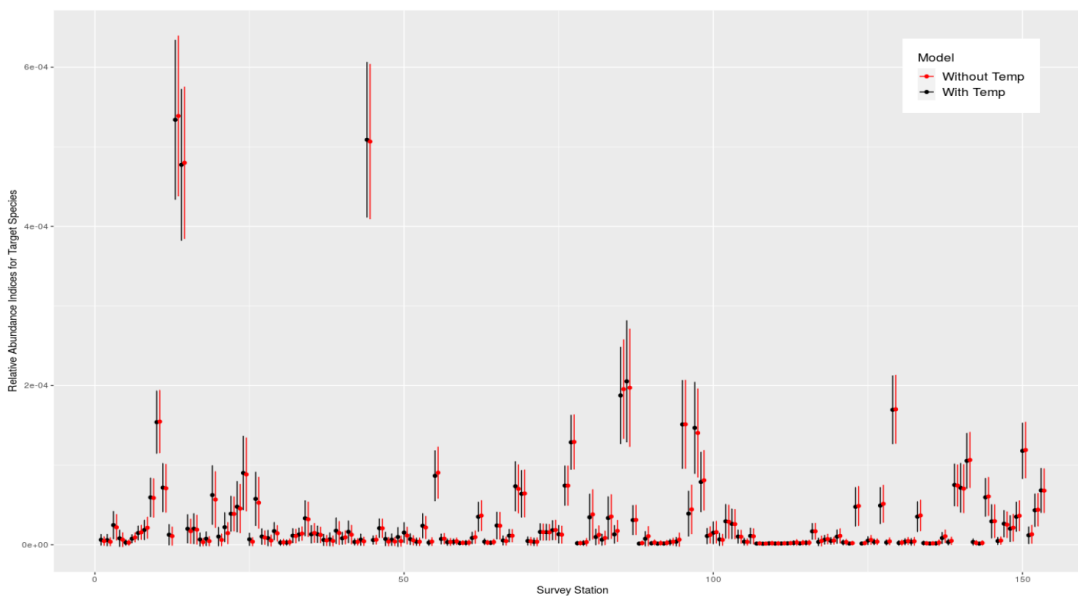


Figure 2.11: Relative abundance indices for halibut and corresponding standard errors estimated by models with and without temperature for 2018.

Chapter 3

Estimated Survey Indices

Here we discuss the relative abundance indices for halibut estimated using our spatial model and how these may be aggregated to obtain annual survey indices.

3.1 Estimated Survey Indices with Corresponding Standard Errors

It can be observed from Figures 3.1 and 3.2 that the estimated survey indices for halibut are higher in the shallow areas. There are also more halibut in the survey area 4X5YZ with depth ranges from 30 to 130 m for 2017, 2018, and 2019. In comparison to 2017, both 2018 and 2019 have more halibut in survey area 4W with depth ranges from 30 to 130 m. Additionally, compared to 2018, both 2017 and 2019 have less halibut in survey area 3P. The highest estimated relative abundance indices of the target species appear in survey areas 4V, 4X5YZ and 4W for 2017, 2018 and 2019 respectively. From Figures 3.3 to 3.5, we can observe that the error bars for larger estimated survey indices are wider.

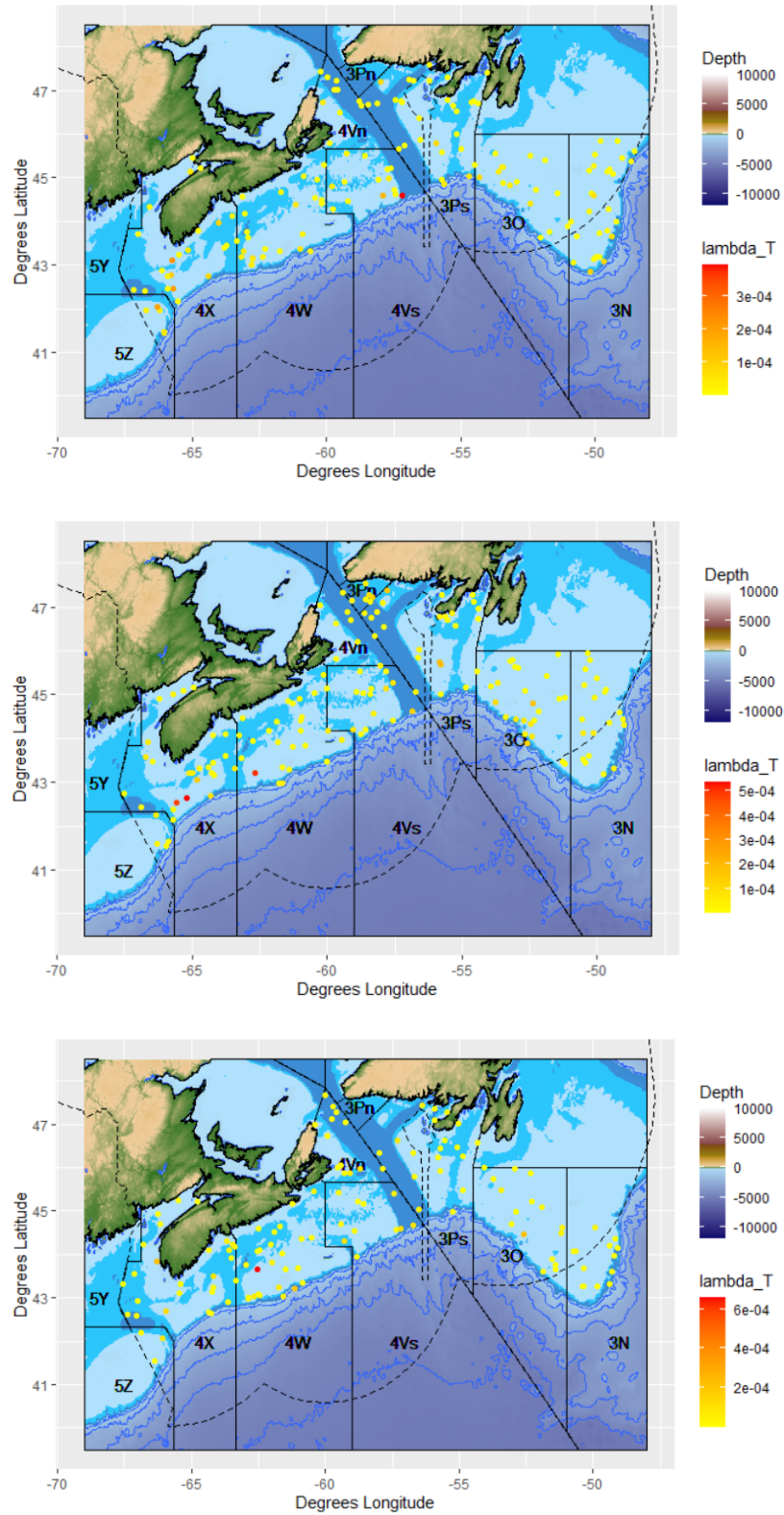


Figure 3.1: Estimated relative abundance indices for halibut for 2017 (top), 2018 (middle) and 2019 (bottom) based on the multinomial exponential model with Gaussian random field.

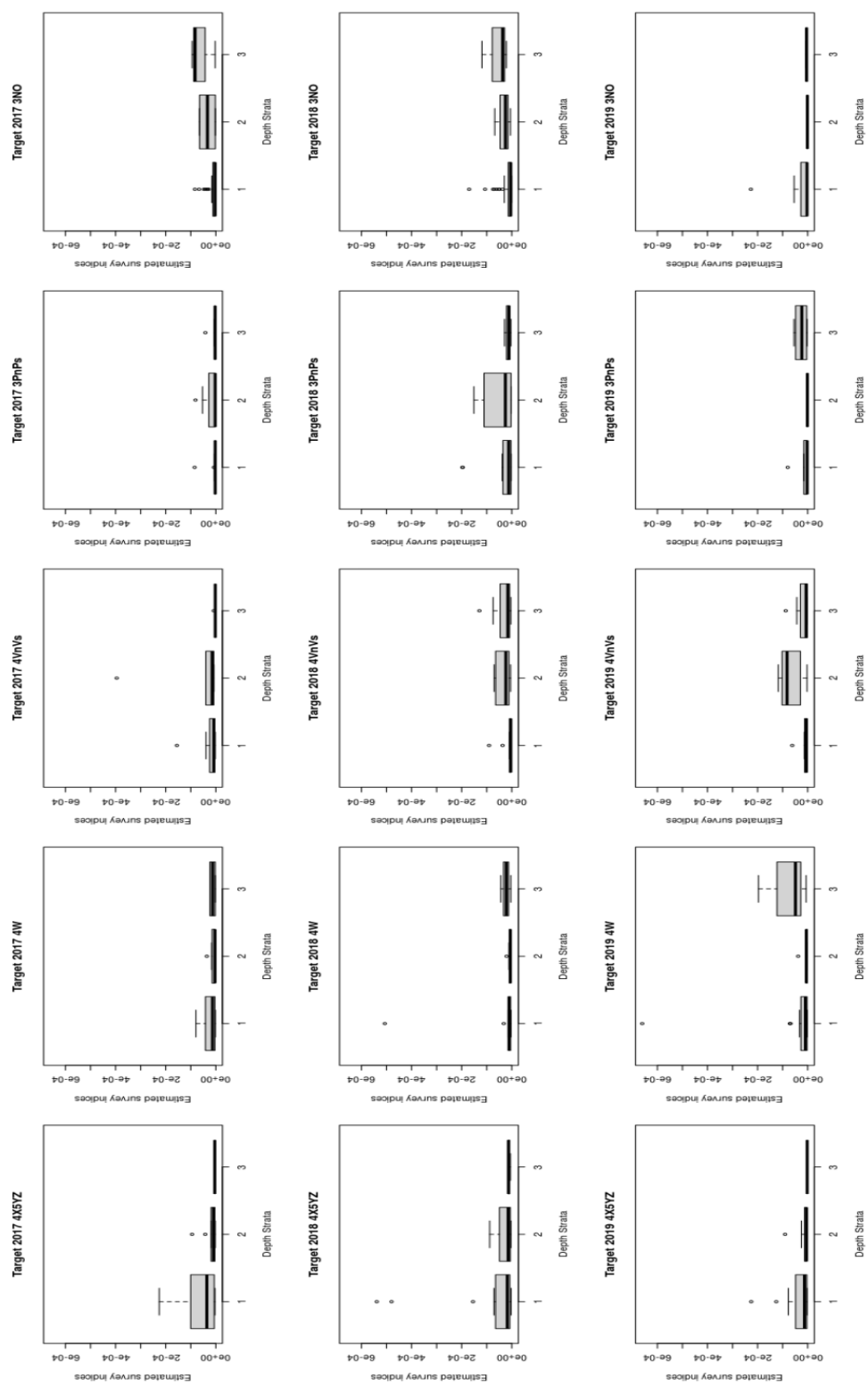


Figure 3.2: Boxplots of classified estimated relative abundance indices based on the multinomial exponential model with Gaussian random field for 2017, 2018 and 2019. The estimates are classified by strata (NAFO Divisions:4X5YZ, 4W, 4V, 3P, 3NO and depth strata: 30-130 m, 131-250 m, 251-750 m).

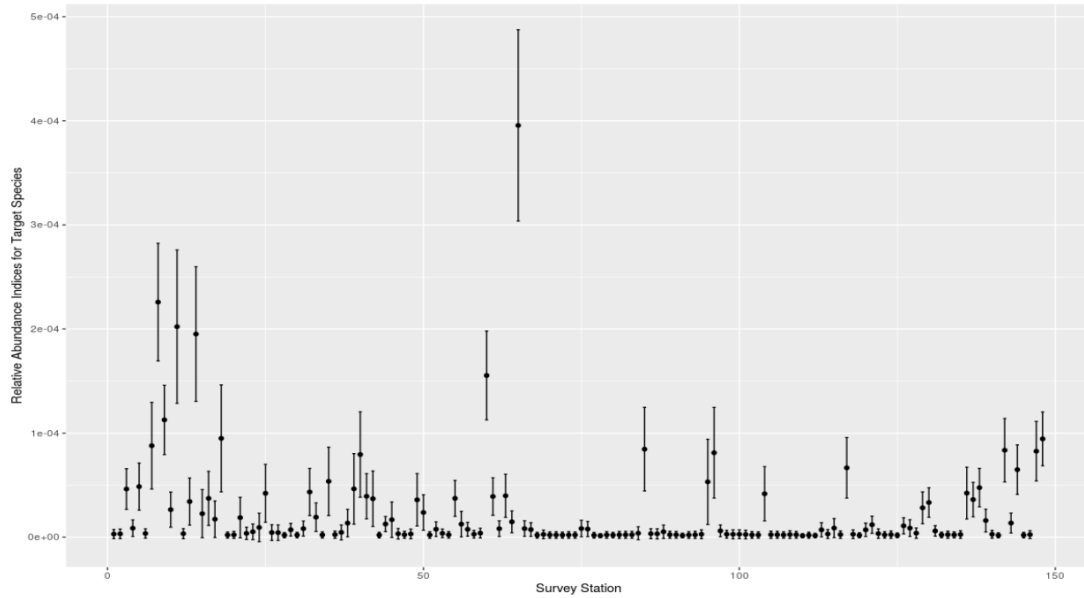


Figure 3.3: Estimated relative abundance indices and corresponding standard errors for halibut for 2017 based on the multinomial exponential model with Gaussian random field.

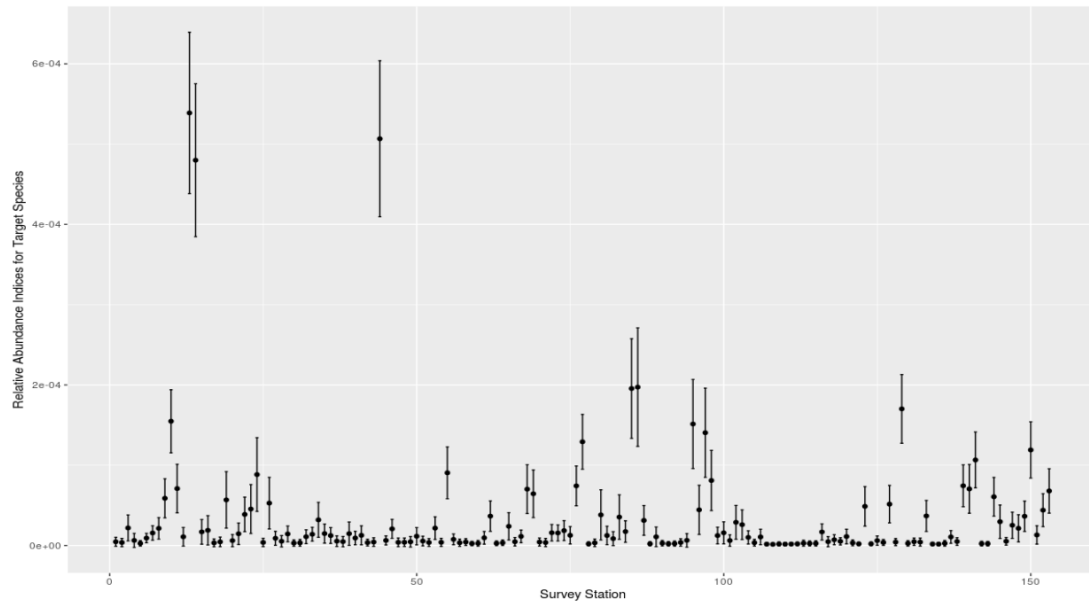


Figure 3.4: Estimated relative abundance indices and corresponding standard errors for halibut for 2018 based on the multinomial exponential model with Gaussian random field.

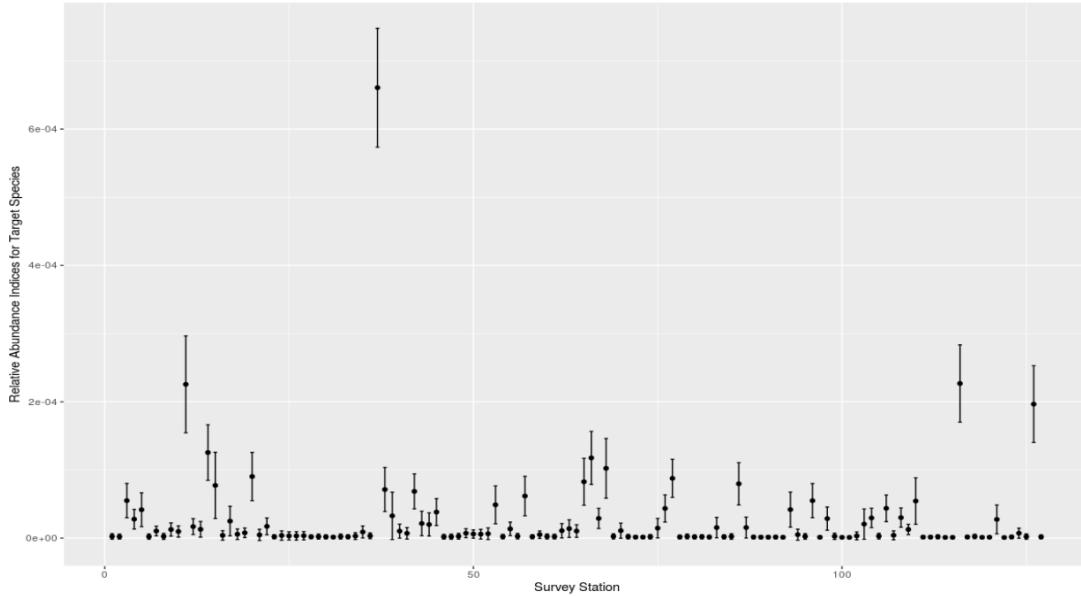


Figure 3.5: Estimated relative abundance indices and corresponding standard errors for halibut for 2019 based on the multinomial exponential model with Gaussian random field.

3.2 Interpolation of Estimated Relative Abundance Indices and Random Field

The spatial model can be used to estimate the survey indices for target and non-target species at the specific survey stations but we would also like to predict the relative abundance indices over the entire survey area, therefore kriging is used. Kriging [16] is a method of interpolation by which the interpolated values are modeled by a Gaussian process governed by prior covariances. The relative abundance indices are interpolated over the block id as the survey locations are selected from the block id. The Gaussian random field ω is interpolated using the formula

$$\hat{Z}(\mathbf{s}^*) = E\{Z(\mathbf{s}^*) \mid \mathbf{Z}\} = \mathbf{B}^\top \mathbf{Z},$$

where

- $\hat{Z}(\mathbf{s}^*)$ are the interpolated Gaussian random field ω at location \mathbf{s}^* which is the location of the block id,
- \mathbf{B} is the cross-covariance matrix between the block id and survey stations,

- $\mathbf{Z} = \Sigma_Z^{-1} \times \bar{Z}$. Σ_Z^{-1} is the inverse of the covariance matrix of the model with estimated spatial parameters (Tables 3.1 and 3.2) for ω based on the location of survey stations. \bar{Z} is the estimated ω at each survey station.

Then the interpolation of the survey indices for target and non-target species can be done using

$$\hat{\lambda}_T(\mathbf{s}^*) = \exp(\beta_{T.0} + \hat{Z}_T(\mathbf{s}^*)) \quad (3.1)$$

$$\hat{\lambda}_{NT}(\mathbf{s}^*) = \exp(\beta_{NT.0} + \hat{Z}_{NT}(\mathbf{s}^*)). \quad (3.2)$$

Figure 3.6 is the interpolated random field and from this interpolation results we can obtain the interpolated survey indices for halibut (Figure 3.7). It can be observed from Figure 3.7 that there more halibut in the survey area 4X5YZ, 4W and 4V.

Year	Variance parameter σ^2 for Target Species	Range Parameter ϕ for Target Species	Smoothness Parameter ν for Target Species
2017	3.528151	0.031142	1
2018	3.162326	0.063860	1
2019	4.399422	0.062808	1

Table 3.1: Estimated parameters of Matérn covariance function for target species for 2017, 2018 and 2019. Smoothness parameter ν is fixed at one.

Year	Variance parameter σ^2 for Target Species	Range Parameter ϕ for Target Species	Smoothness Parameter ν for Target Species
2017	1.538283	0.086193	1
2018	1.033545	0.06819973	1
2019	1.495082	0.05448649	1

Table 3.2: Estimated parameters of Matérn covariance function for non-target species for 2017, 2018 and 2019. Smoothness parameter ν is fixed at one.

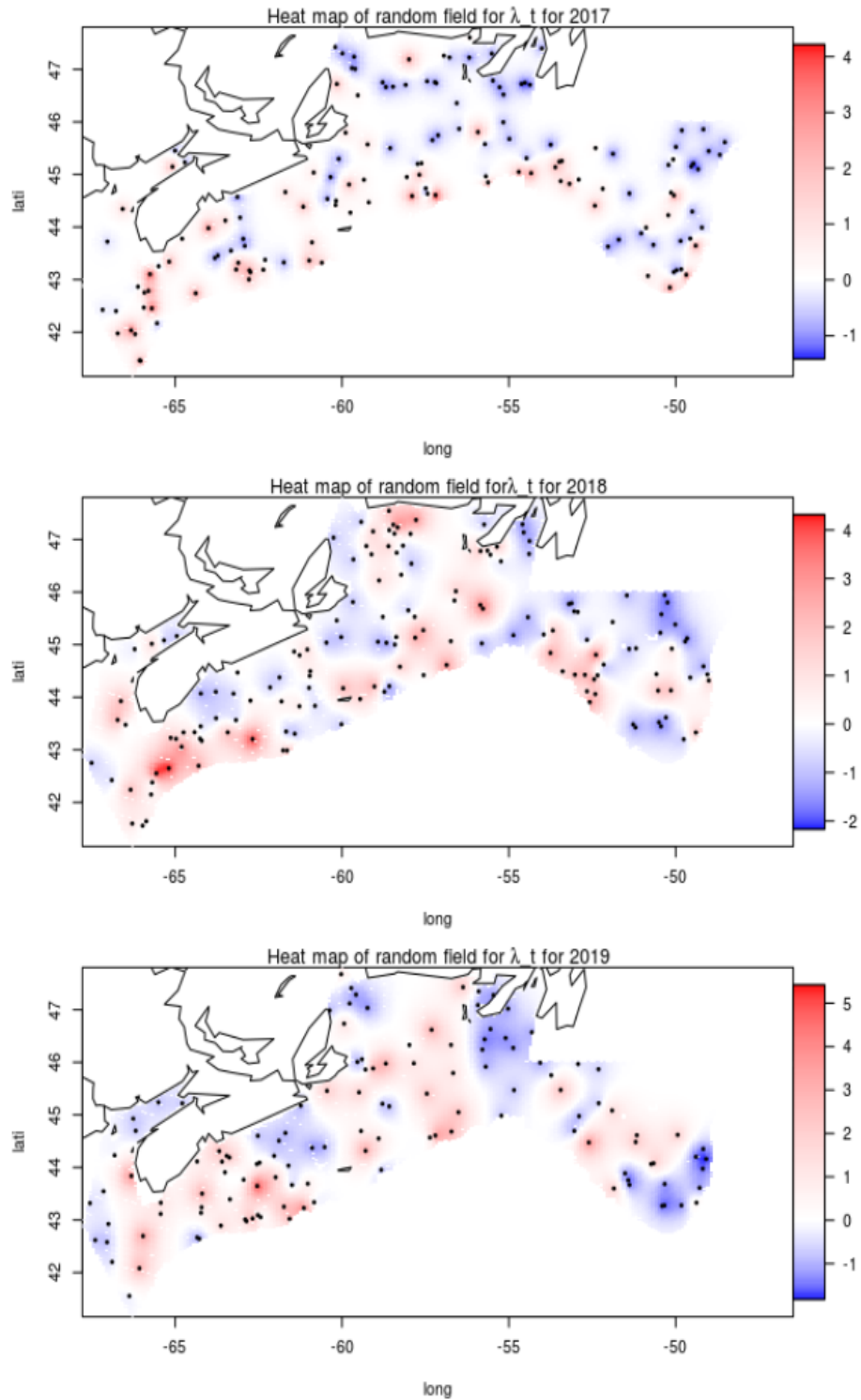


Figure 3.6: Interpolated Gaussian random fields for halibut for 2017 (top), 2018 (middle) and 2019 (bottom). The black points are locations of the survey stations for each year.

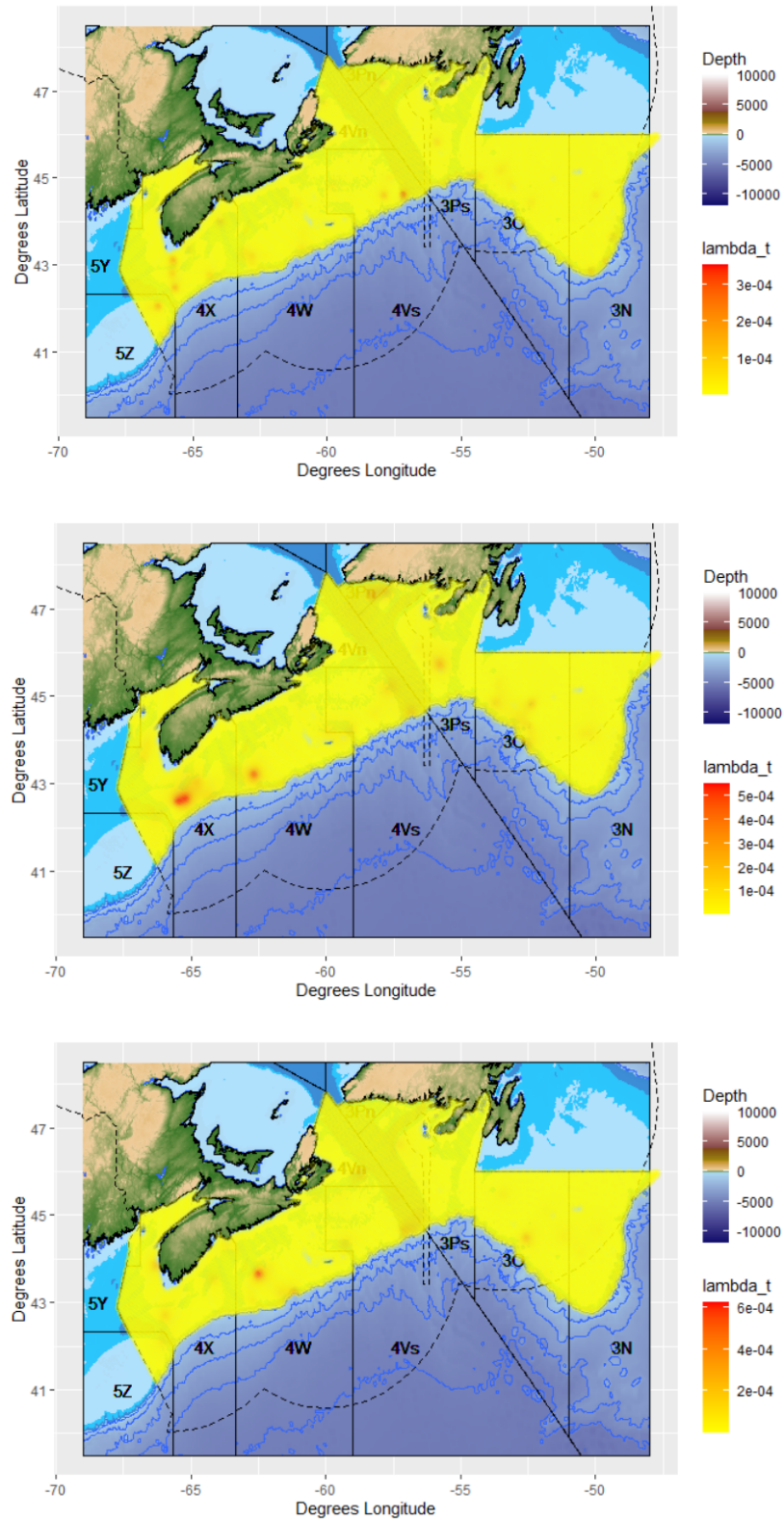


Figure 3.7: The heatmaps of the interpolated relative abundance indices of halibut for 2017 (top), 2018 (middle) and 2019 (bottom).

3.3 Weighted Average of the Survey Indices

The aggregation of survey indices and the corresponding standard errors have been estimated using three different approaches so as to compare the changes in relative abundance indices of halibut over the past three years.

3.3.1 Explicit Method and corresponding Standard Error

The explicit method uses Equations 1.1 and 1.2 of the previous section to calculate the relative abundance indices for target and non-target species within each stratum. Then the weighted average for the entire survey area is obtained using the sum of the indices weighted by the corresponding stratum area and divided by the sum of the stratum area (Equation 3.1). In total, there are 15 strata with 3 depth strata and 5 area strata. However, stratum H13 is excluded from this analysis as the number of target species caught in 2017 for stratum H13 was 0 which would cause errors when calculating the survey indices. Bootstrapping is used to estimate the standard error of the estimated area weighted average of the survey indices [12]. Specifically, function `bootstrap` from R package `bootstrap` [25] is used to generate 15000 leave-one-out bootstrap replicates of Equations 1.1 and 1.2 applied to survey data. Then we calculate the standard error of observed value of Equations 1.1 and 1.2 applied to survey data.

$$\text{Area Weighted Average} = \frac{\sum_{i=1}^n A_i \lambda_i}{\sum_{i=1}^n A_i} \quad (3.3)$$

where

- A_i is the i^{th} area of interest,
- λ_i is the corresponding estimated relative abundance index for the i_{th} area,
- n is the total number of area of interest.

3.3.2 Mean of the Interpolated Survey Indices and the Corresponding Standard Error

The aggregation of the survey indices via the interpolation method is over 1000 randomly selected locations from the block id. We use the built-in kriging function from

R package TMB to interpolate relative abundance indices over the 1000 randomly selected locations and then calculate the mean of the interpolated results. The standard error for this method is approximated using the delta method [6]:

let $\widehat{\boldsymbol{\theta}}_n$ be a sequence of $K \times 1$ random vectors such that

$$\sqrt{n} \left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} \right) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma})$$

where $N(\mathbf{0}, \boldsymbol{\Sigma})$ is a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$, $\boldsymbol{\theta}$ is a constant $K \times 1$ vector, and \xrightarrow{d} indicates convergence in distribution. Let $\mathbf{g} : \mathbb{R}^K \rightarrow \mathbb{R}^L$. If all the L entries of \mathbf{g} have continuous partial derivatives with respect to $\boldsymbol{\theta}$, then

$$\sqrt{n} \left(\mathbf{g} \left(\widehat{\boldsymbol{\theta}}_n \right) - \mathbf{g}(\boldsymbol{\theta}) \right) \xrightarrow{d} N(\mathbf{0}, \nabla \mathbf{g}(\boldsymbol{\theta}) \boldsymbol{\Sigma} \nabla \mathbf{g}(\boldsymbol{\theta})^\top)$$

where $\nabla \mathbf{g}(\boldsymbol{\theta})$ is the $L \times K$ matrix of partial derivatives of the entries of \mathbf{g} with respect to the entries of $\boldsymbol{\theta}$.

3.3.3 Dirichlet Method and the Corresponding Standard Error

The Dirichlet tessellation method aggregates the survey indices at the level of the longline set. The Dirichlet tessellation of the survey stations is computed using the functions from the R package spatstat [4]. To be more specific, we use the function ppp from the R package spatstat and the locations of survey stations to form the spatial point pattern and specify the enclosing polygon with the boundary points of survey area. The Dirichlet tile associated with a particular point i is the region of space that is closer to point i than to any other point in the spatial point pattern. Then we use function dirichlet from the R package spatstat to divide the survey area with Dirichlet tiles into disjoint regions, forming the tessellation (Figure 3.8). Finally we use the area of the Dirichlet tessellation and the corresponding estimated survey indices to calculate the weighted average (Equation 3.1). The delta method is used to estimate the standard error [6].

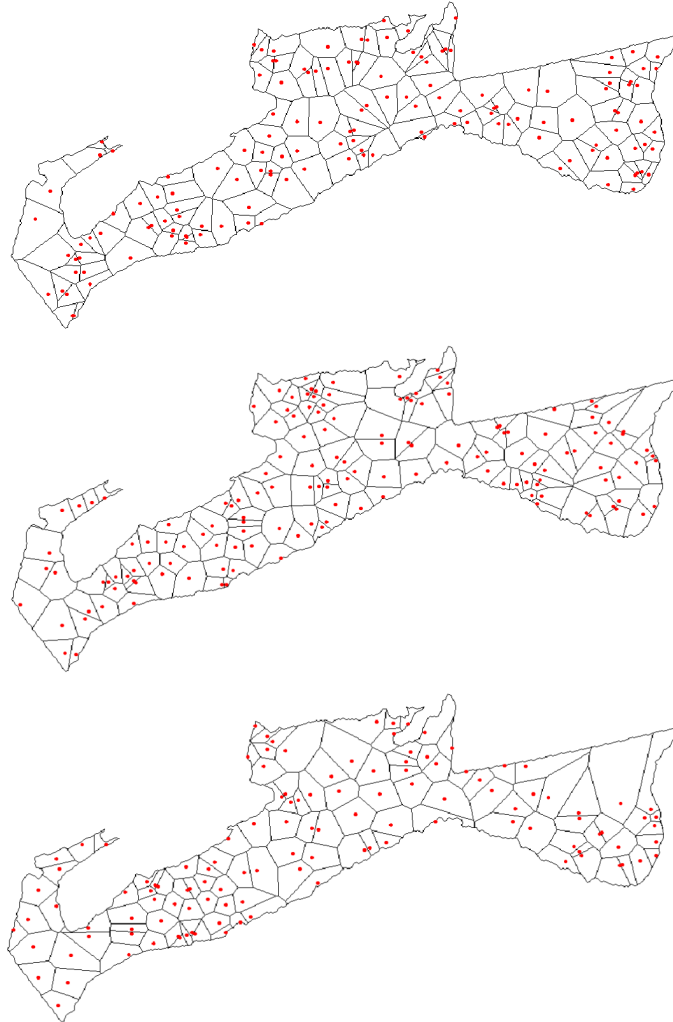


Figure 3.8: The Dirichlet tessellation and the survey stations for 2017 (top), 2018 (middle) and 2019 (bottom).

It can be observed from Figure 3.9 that the weighted relative abundance indices for halibut obtained using these three methods show a similar trend. The weighted survey indices increase from 2017 to 2018 and decrease from 2018 to 2019. The weighted survey indices obtained using the Dirichlet tessellation method are the largest when compared to the other two methods. Standard errors of Dirichlet tessellation method are the smallest among these three methods.

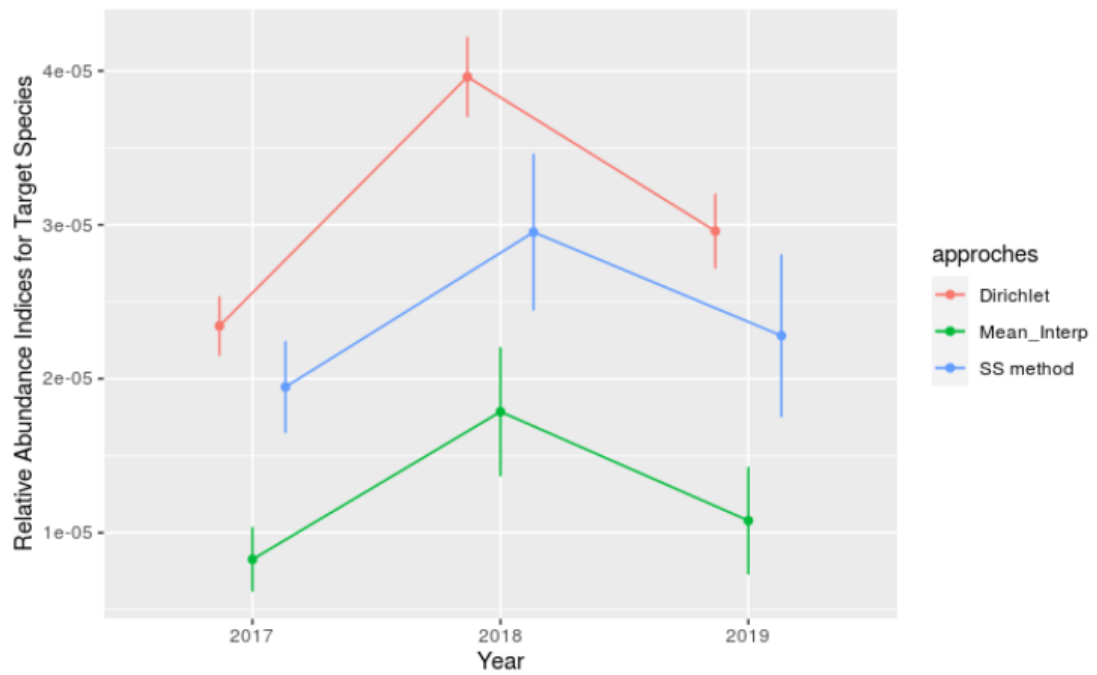


Figure 3.9: The weighted survey indices and corresponding standard errors for halibut obtained using three different weighting methods (all strata except H13) for 2017, 2018 and 2019.

Chapter 4

Discussion

Our results demonstrate (see Figures 3.1, 3.2 and 3.7) that catch rates for halibut are higher in shallow areas in southwest Nova Scotia (relative to shallow areas on Southern Grand Banks) and along shelf edges throughout the management unit. These conclusions are consistent with previous studies of Atlantic Halibut distribution. More recent presentations of commercial landings [11][15] also show similar patterns. Furthermore halibut catch rates in the DFO research vessel trawl surveys [8][15] showed juvenile hotspots and preferred habitat for halibut on shallow banks in southwest Nova Scotia (NAFO 4X and 4W, Strata 1) and along the shelf edge throughout the management unit.

We have shown that the differences between estimated survey indices for halibut obtained using fixed soak time and set-specific soak time are quite small when the explicit model is used. However, these differences are larger when the indices are estimated at the level of the longline set with our spatial model. It can be observed from Figures 4.1 through 4.3 that at a few survey stations, where the largest indices occur, the indices estimated using constant soak time are much smaller than the indices obtained using actual soak time. It also can be observed from these three plots that the estimates with actual soak time and mean soak time differ more when the estimated survey indices are large. Further investigation could be conducted to examine the causing of these large differences.

Although there are 1000 hooks sampled on a longline set, hook conditions are only observed for 300 of these hooks. This information is then used to estimate the effort of a longline set. For the remaining 700 hooks, the total number of halibut and non-target species caught are recorded. Earlier simulation analyses [24] suggest that 300 hooks are representative but the relative abundance indices for halibut estimated

using only this information are low. This suggests that the additional information from the remaining 700 hooks could be included in future analyses to obtain more precise estimates.

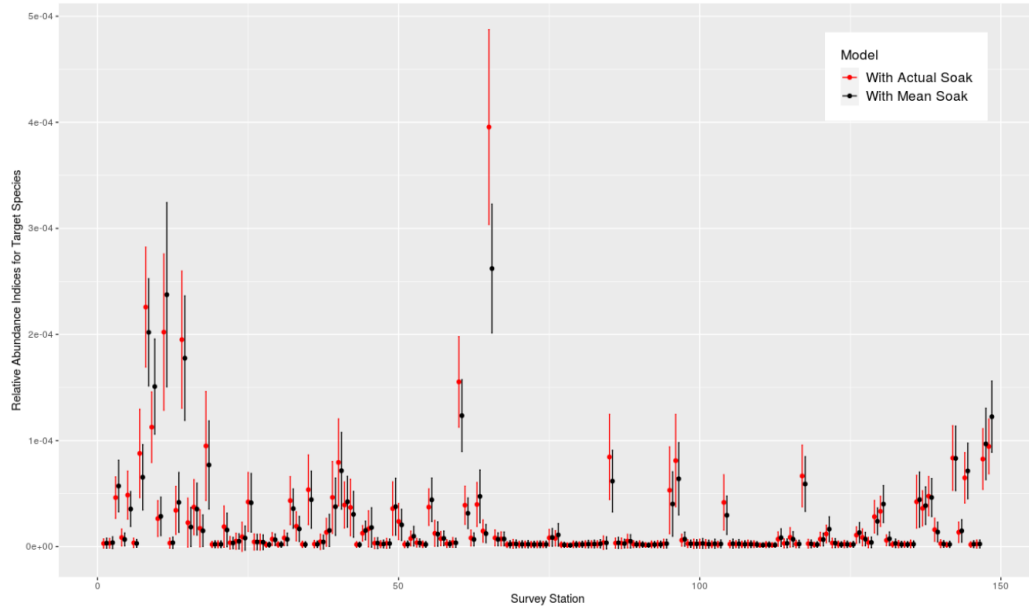


Figure 4.1: Relative abundance indices for halibut and corresponding standard errors estimated by models with set-specific soak time and mean soak time for 2017.

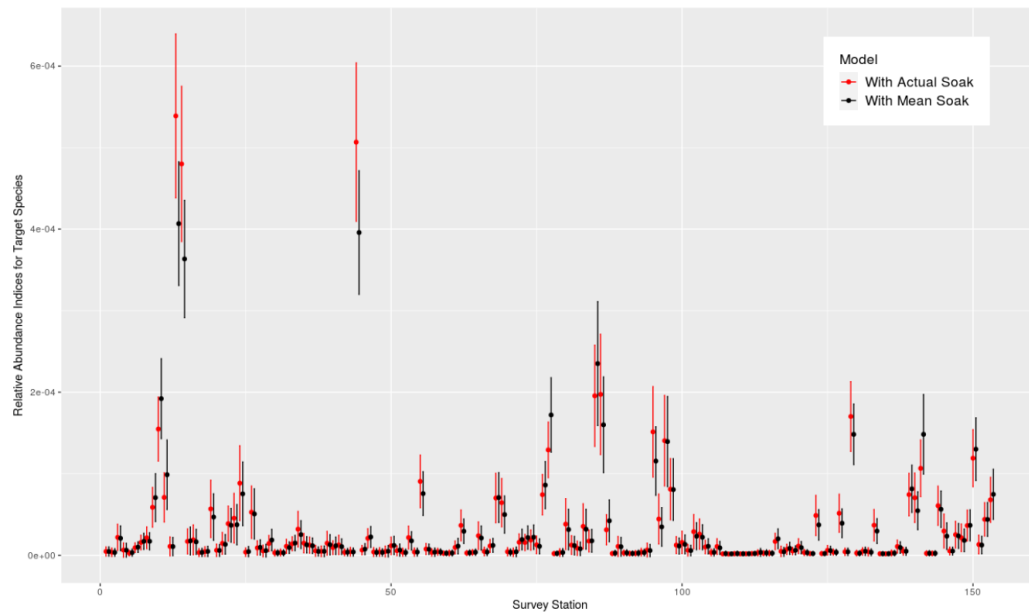


Figure 4.2: Relative abundance indices for halibut and corresponding standard errors estimated by models with set-specific soak time and mean soak time for 2018.

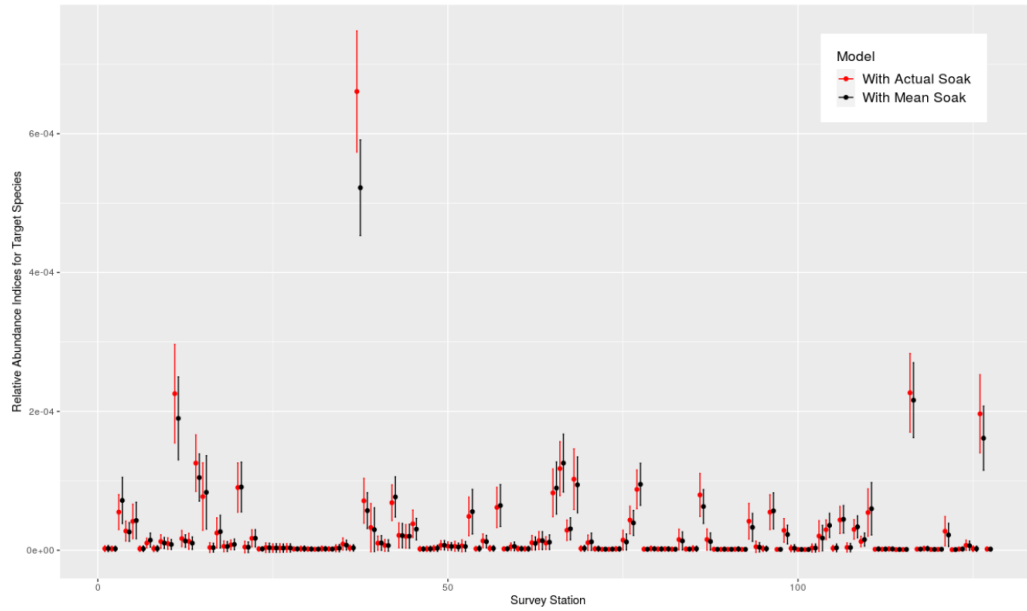


Figure 4.3: Relative abundance indices for halibut and corresponding standard errors estimated by models with set-specific soak time and mean soak time for 2019.

Bibliography

- [1] MSC 2nd Annual Surveillance Audit Report For The Canadian Atlantic Halibut Fishery Facilitated By The Atlantic Halibut Council. SAI Global Assurance Services, 2015.
- [2] A large endangered Atlantic Halibut (*Hippoglossus hippoglossus*) seen resting on the seafloor during Dive 05 of the Deep Connections 2019 expedition. <https://oceanexplorer.noaa.gov/oceanos/explorations/ex1905/dailyupdates/sept2/sept2.html>, 2019 (accessed January 11, 2020).
- [3] R. J. Adler and J. Taylor. *Random Fields and Geometry*. Springer, New York, 2007.
- [4] A. Baddeley, E. Rubak, and R. Turner. *Spatial Point Patterns: Methodology and Applications with R*. Chapman and Hall/CRC Press, London, 2015. Available at <http://www.crcpress.com/Spatial-Point-Patterns-Methodology-and-Applications-with-R/Baddeley-Rubak-Turner/9781482210200/>.
- [5] R.J.H. Beverton and S.J. Holt. *On the dynamics of exploited fish populations*. Her Majesty’s Stationary Office, London, 533 p, 1957.
- [6] P.J. Bickel and K.A. Doksum. *Mathematical Statistics: Basic Ideas and Selected Topics, Volumes I*. Chapman and Hall/CRC Press, 2015. doi:10.1201/9781315369266.
- [7] Å. Bjordal. Effect of different long-line baits (mackerel, squid) on catch rates and selectivity for tusk and ling. *ICES C.M.*, 1983/B:31:9 p, 1983.
- [8] S.A. Boudreau, S. Shackell, N.L. Carson, and C.E. den Heyer. Connectivity, persistence, and loss of high abundance areas of a recovering marine fish population in the Northwest Atlantic Ocean. *Ecol. Evol.*, 7:9739–9749, 2017.
- [9] M.E. Brooks, K. Kristensen, K.J. van Benthem, A. Magnusson, C.W. Berg, A. Nielsen, H.J. Skaug, M. Maechler, and B.M. Bolker. glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal*, 9(2):378–400, 2017. Available at <https://journal.r-project.org/archive/2017/RJ-2017-066/index.html>.
- [10] S.P. Cox, A. Benson, and C.E den Heyer. Framework stock assessment for Atlantic halibut on the Scotian Shelf and Southern Grand Banks. *DFO Can. Sci. Advis. Sec. Res. Doc.*, 2016/001:v + 54 p, 2018.

- [11] C. den Heyer, B. Hubley, D. Themelis, S.C. Smith, S. Wilson, and G. Wilson. Atlantic halibut on the Scotian Shelf and Southern Grand Banks: Data review and assessment model update. *DFO Can. Sci. Advis. Sec. Res. Doc.*, 2015/051:v + 82 p, 2015.
- [12] B. Efron and R.J. Tibshirani. *An introduction to the bootstrap*. Chapman and Hall, London, 437 p, 1993.
- [13] M.P. Etienne, S.G. Obradovich, K.L. Yamanaka, and M.K. McAllister. Extracting abundance indices from longline surveys: a method to account for hook competition and unbaited hooks. *arXiv 1005.0892v3*, pages 1–35, 2013.
- [14] E. Fernandez and L.M. Lellouche. *Product user manual for the global ocean physical reanalysis product GLOBAL_REANALYSIS_PHY_001_030*. Issue 1.1. Copernicus Marine Environment Monitoring Service., EU, 2018. Available at <https://resources.marine.copernicus.eu/documents/PUM/CMEMS-GLO-PUM-001-030.pdf>.
- [15] K. French, N.L. Shackell, and C.E. den Heyer. Habitat suitability of Atlantic halibut (*Hippoglossus hippoglossus*) in the Northwest Atlantic. *Fishery Bulletin*, 116(2):111–125, 2017.
- [16] A.E. Gelfand, M. Fuentes, P. Guttorp, and P. Diggle. *Handbook of Spatial Statistics*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. Taylor & Francis, 2010. Available at <http://books.google.com/books?id=EFbbcMFZ2mMC>.
- [17] H. Hovgård and H. Lassen. Manual on estimation of selectivity for gillnet and longline gears in abundance surveys. *FAO Fish. Tech. Paper. FAO, Rome.*, 397:84 p, 2000.
- [18] K. Kristensen, B. Bell, H. Skaug, A. Magnusson, C. Berg, A. Nielsen, M. Maechler, T. Michelot, M. Brooks, A. Forrence, C.M. Albertsen, and C. Monnahan. *Template Model Builder: A General Random Effect Tool Inspired by ADMB*, 2020. Available at <https://cran.r-project.org/web/packages/TMB/TMB.pdf>.
- [19] K. Kristensen, A. Nielsen, C.W. Berg, H. Skaug, and B.M. Bell. TMB: Automatic Differentiation and Laplace Approximation. *Journal of Statistical Software*, 70(5):1–21, 2016. doi:10.18637/jss.v070.i05.
- [20] E. Pante and B. Simon-Bouhet. marmap: A Package for Importing, Plotting and Analyzing Bathymetric and Topographic Data in R. *PLoS ONE*, 8(9):e73051, 2013. doi:10.1371/journal.pone.0073051.
- [21] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. Available at <https://www.R-project.org/>.

- [22] B.J Rothschild. Competition for gear in a multiple-species fishery. *J. Cons. perm. int. Explor. Mer*, 31:102—110, 1967.
- [23] S. Smith. *Analysis of longline survey data*, 2016. This package fits the multinomial model to longline survey data discussed in Smith (2016).
- [24] S. Smith. Review of the Atlantic halibut longline survey index of exploitable biomass. *Canadian Technical Report of Fisheries and Aquatic Sciences*, 3180:v + 56 p, 2016.
- [25] R. Tibshirani and F. Leisch. *Functions for the Book “An Introduction to the Bootstrap”*, 2019. Available at <https://cran.r-project.org/web/packages/bootstrap/bootstrap.pdf>.
- [26] M.K. Trzcinski, S.L. Armsworthy, S. Wilson, R.K. Mohn, M. Fowler, and S.E. Campana. Atlantic halibut on the Scotian Shelf and Southern Grand Banks (NAFO Divisions 3NOPs4VWX5Zc) — Industry/DFO longline survey and tagging results to 2008. *DFO Can. Sci. Advis. Sec. Res. Doc.*, 2009/026:vi + 43 p, 2009.
- [27] R.A. Webster and S.R. Hare. Adjusting IPHC setline survey WPUE for survey timing, hook competition and station depth. *IPHC Rep. Assess. Res. Act.*, 2009:187—208, 2009.
- [28] H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. Available at <https://ggplot2.tidyverse.org>.
- [29] K.C.T. Zwanenburg and S. Wilson. The Scotian Shelf and Southern Grand Banks Atlantic halibut (*Hippoglossus hippoglossus*) survey — Collaboration between the fishing and fisheries science communities. *ICES CM*, 2000/W:20:25p, 2000.

Appendix

A.1 Survey Indices Reproduction

The following codes are for reproducing the llsurv survey indices with the multinomial exponential model (MEM1) and R package TMB. The model is formulated in C++ and the data are manipulated in R.

- C++ Code for Model Formulation:

```

1 #include <TMB.hpp> // Links in the TMB libraries
2 template<class Type>
3 Type objective_function<Type>::operator() ()
4 {
5     DATA_MATRIX(x); // The halibut data
6     PARAMETER_VECTOR(theta); // Parameters
7     DATA_VECTOR(s); // Soak time
8     int n = x.col(0).size(); // Length of the column of matrix x
9     int n_k = x.row(0).size(); // Length of the row of matrix x (4)
10
11     // ldat - the relative abundance index for target species (halibut
12     )
13     vector<Type> ldat(n);
14     for(int i=0; i<n; ++i){
15         ldat(i)=exp(theta(0));
16     }
17     ADREPORT(ldat);
18
19     // ldant - the relative abundance index for non-target species
20     vector<Type> ldant(n);
21     for(int i=0; i<n; ++i){
22         ldant(i)=exp(theta(1));
23     }
24     ADREPORT(ldant);
25
26     // pnt - the escaping rate for non-target species
27     vector<Type> pnt(n);

```

```

27  for(int i=0; i<n; ++i){
28      pnt(i)=invlogit(theta(2));
29  }
30  ADREPORT(pnt);
31
32  matrix<Type> p(n, n_k);
33  for(int i=0;i<n;i++){
34      p(i,0) = exp(-(ldat(i)+ldant(i))*s(i));
35      p(i,1)=(Type(1)-exp(-(ldat(i)+ldant(i))*s(i)))*(ldat(i)/(ldat(i)
+ldant(i)));
36      p(i,2)=(Type(1)-exp(-(ldat(i)+ldant(i))*s(i)))*(ldant(i)/(ldat(i)
+ldant(i)))*(Type(1)-pnt(i));
37      p(i,3)=1-p(i,0)-p(i,1)-p(i,2);
38  }
39
40  Type nll;
41  for(int i=0; i < n; i++){
42      vector<Type> p_row=p.row(i);
43      vector<Type> x_row=x.row(i);
44      nll -= dmultinom(x_row,p_row,true);
45  }
46
47  REPORT(ldat);
48  REPORT(ldant);
49  REPORT(pnt);
50  return nll;
51 }

```

- R Code for Data Manipulation:

```

1  ##For 2017
2  # Clear memory
3  rm(list=ls())
4  # The halibut data for 2017
5  data_2017 = read.csv("hdata2017.csv", header = T, stringsAsFactors =
F)
6  head(data_2017)

```

```

7 dim(data_2017)
8 # Corrected Strata ID
9 st_id = read.csv("HS_STATION_STRATA_2017_18_19.csv", header = F,
  stringsAsFactors = F)
10 names(st_id) = c("YEAR", "STATION", "STRATA")
11 st_id_17 = st_id[which(st_id$YEAR==2017), ]
12 # Merge the data
13 data2017 = merge(data_2017, st_id_17, by = "STATION")
14
15 # Drop the stations (523 and 525) with hook = 30
16 not30 = which(data2017$hooks_sampled == 30)
17 data2017[not30,]
18 data2017 = data2017[-c(not30), ]
19 which(data2017$hooks_sampled == 30)
20
21 ## Data
22 # st is the soak time
23 # nb is the number of baited hooks at the end of the soak time
24 # nt is the number of individuals of the target species caught
25 # nnt is the number of individuals of the non-target species caught
26 # ne is the number of empty hooks at the end of the soak time
27 nb_17=data2017$empty_baited
28 nt_17=data2017$halibut
29 nnt_17=data2017$other_species
30 ne_17=data2017$empty_unbaited+data2017$broken_hook
31 st_17=data2017$SOAKMINP3P1
32 # The mean of the soak time (fix at the mean of the soak time)
33 st_17fixed=rep(mean(st_17),length(st_17))
34
35 ## Calculate the initial values for TMB estimation
36 # Using the explicit equation to calculate the estimators (MEM1 with
  pt=0 from Etienne 2010)
37 soak_17=mean(st_17)
38 # lambda for target species

```

```

39 ldathat_17=(sum(nt_17)/(sum(nt_17+nnt_17+ne_17+nb_17)-sum(nb_17)))*
    (1/soak_17)*log(sum(nt_17+nnt_17+ne_17+nb_17)/sum(nb_17))
40 ldathat_17
41 # lambda non-target species
42 ldanthat_17=((sum(nnt_17)+sum(ne_17))/(sum(nt_17+nnt_17+ne_17+nb_17)
    -sum(nb_17)))*(1/soak_17)*log(sum(nt_17+nnt_17+ne_17+nb_17)/sum(
    nb_17))
43 ldanthat_17
44 # Escaping probabilitly for target species
45 pthat_17=0
46 pthat_17
47 # Escaping probability for non-target species
48 pnthat_17=sum(ne_17)/(sum(ne_17)+sum(nnt_17))
49 pnthat_17
50
51 # Logit function
52 logitp=function(p){log(p/(1-p))}
53 # Inverse logist function
54 logitpi=function(t){exp(t)/(1+exp(t))}
55
56 ## Use TMB to reproduce the results
57 library(TMB)
58 # Call TMB function value
59 compile("rep_code.cpp")
60 # Dynamically link the C++ code
61 dyn.load(dynlib("rep_code"))
62
63 data2017$empty = data2017$empty_unbaited+data2017$broken_hook
64 # Put the required data into a matrix
65 x_17 = with(data2017, cbind(empty_baited, halibut, other_species,
    empty))
66 # Soak time
67 st_17 = as.vector(st_17)
68 st_17fixed = as.vector(st_17fixed)
69

```



```

70 # Data list
71 data_17 = list(x=x_17,s=st_17)
72 data_17fixed = list(x=x_17,s=st_17fixed)
73
74 # Parameter list
75 param_17 = list()
76 param_17fixed = list()
77 # Initial values for lambda.t, lambda.nt and pnt
78 # Use the values calculated previously as the starting points
79 param_17$theta = c(log(ldathat_17),log(ldanthat_17),logitp(pnthat_
      17))
80 param_17fixed$theta = c(log(ldathat_17),log(ldanthat_17),logitp(
      pnthat_17))
81
82 # Construct an R object (f) that represents our C++ function
83 # MakeADFun calls C++
84 f_17 = MakeADFun(data_17, param_17, DLL="rep_code")
85 f_17fixed = MakeADFun(data_17fixed, param_17fixed, DLL="rep_code")
86
87 # Call TMB function value
88 fit_17 = nlminb(f_17$par,f_17$fn,f_17$gr)
89 fit_17fixed = nlminb(f_17fixed$par,f_17fixed$fn,f_17fixed$gr)
90 # Calculate standard deviations of all model parameters
91 sdr_17 = sdreport(f_17)
92 # Estimated results for model with actual soak time
93 summary(sdr_17)
94 sdr_17fixed = sdreport(f_17fixed)
95 # Estimated results for model with actual soak time
96 summary(sdr_17fixed)

```

A.2 Binomial Model with Gaussian Random Field (glmmTMB)

The following codes are used for estimating the probability of getting unbaited hooks over 1000 randomly selected survey stations from block id. R package glmmTMB and binomial model with Gaussian random field are used.

```
1 ## For 2017
2 # Clear memory
3 rm(list=ls())
4 library(ggplot2)
5 library(dplyr)
6 library(marmap)
7
8 # The halibut data for 2017
9 data_2017 = read.csv("hdata2017.csv", header = T, stringsAsFactors =
    F)
10 head(data_2017)
11 dim(data_2017)
12
13 # Corrected Strata ID
14 st_id = read.csv("HS_STATION_STRATA_2017_18_19.csv", header = F,
    stringsAsFactors = F)
15 names(st_id) = c("YEAR", "STATION", "STRATA")
16 st_id_17 = st_id[which(st_id$YEAR==2017), ]
17 # Merge the data
18 data2017 = merge(data_2017, st_id_17, by = "STATION")
19 head(data2017)
20 dim(data2017)
21
22 # Drop the stations (523 and 525) with hook = 30
23 data2017 = data2017[-c(not30), ]
24 which(data2017$hooks_sampled == 30)
25 # Read in the block id for the prediction
26 blockid = read.csv("blockIDkey.csv", header = T)
27 blockid[1:4, ]
28
29 # For binomial, we have baited and unbaited hooks (unbaited hooks +
    halibut + other species
30 # + broken hooks)
31 # Calculate the proportion unbaited hooks
```

```

32 data2017$biprop_unbaited = (data2017$hooks_sampled-data2017$empty_
    baited)/data2017$hooks_sampled
33 data2017$biprop_unbaited[1:10]
34
35 data17 = data2017 %>% dplyr::select(P1LONG.x, P1LAT.x, P1DEPTH,
    empty_baited, hooks_sampled, SOAKMINP3P1, MED_TEMP)
36 data17$biunbaited=data17$hooks_sampled-data17$empty_baited
37 d17 = data.frame(x = as.vector(data17$P1LAT.x),
38                 y = as.vector(data17$P1LONG.x))
39
40 library(glmTMB)
41 # To fit the model, a numFactor and a dummy grouping variable must
    be added to the dataset:
42 d17$pos <- numFactor(d17$x, d17$y)
43 d17$group <- factor(rep(1, nrow(d17)))
44 d17$soak = data2017$SOAKMINP3P1
45 d17$unbaited = data17$biunbaited
46 d17$baited = data17$empty_baited
47 d17$depth = data17$P1DEPTH
48
49 # With covariate temperature
50 mat_17 <- glmTMB(cbind(unbaited,baited) ~ 1 + soak + depth + mat(
    pos + 0 | group),
51                 family=binomial(link="logit"), data=d17)
52 summary(mat_17)
53
54 # Random sample 1000 points from the the block id without
    replacement for prediction
55 samp = blockid[sample(nrow(blockid),size=1000,replace=FALSE),]
56 idpos = numFactor(samp$lat.DecDeg, samp$lon.DecDeg)
57 # Format the newdata which will be used for prediction
58 newdata_17 <- data.frame(pos=idpos, soak = mean(data2017$SOAKMINP3P1
    ), depth = samp$depth_m)
59 newdata_17$group <- factor(rep(1, nrow(newdata_17)))
60 head(newdata_17)

```

```

61
62 # Predict the proportion of unbaited hooks
63 p_17=predict(mat_17, newdata_17, type="response", allow.new.levels=
      TRUE)
64 head(p_17)
65
66 # Reshape the predicted proportion and corresponding lat and lon to
      a data.frame
67 matn_17=data.frame(lati=samp$lat.DecDeg,
68                   long=samp$lon.DecDeg,
69                   p=p_17)
70 write.csv(matn_17, file="matn_17.csv")
71
72 # Plot the predicted proportion of unbaited hooks
73 ggplot(matn_17, aes(x = long, y = lati, colour = p)) + geom_point()
      +
74   scale_colour_gradient(low = "yellow", high = "red")

```

A.3 Multinomial Exponential Model (MEM1) with Gaussian Random Field without Covariate (TMB)

The following codes are used for estimating the relative abundance indices for target and non-target species at each survey station with spatial model without covariates and stratum H13.

- C++ Code for Model Formulation:

```

1 #include <TMB.hpp> // Links in the TMB libraries
2 template<class Type>
3 Type objective_function<Type>::operator() ()
4 {
5   DATA_MATRIX(H); // The halibut data
6   DATA_MATRIX(X); // Covariates matrix
7   PARAMETER_VECTOR(betat); // Parameters for lambda t (target
      species)

```

```

8  PARAMETER_VECTOR(betant); // Parameters for lambda nt (non-target
   species)
9  PARAMETER(theta); // Parameters for pnt (probability of caught non
   -target fish escaping)
10 DATA_VECTOR(s); // Soak time
11 int n = H.col(0).size(); // Length of the column of matrix H
12 int n_k = H.row(0).size(); // Length of the row of matrix H (4)
13 PARAMETER_VECTOR(omegat); // The random field for lambda t
14 PARAMETER_VECTOR(omegant); // The random field for lambda nt
15
16 // ldat - the relative abundance indices for target species (
   halibut)
17 vector<Type> ldat(n);
18 for(int i=0; i<n; ++i){
19     // Adding random field and covariates to lambda t
20     ldat(i)=exp((vector<Type>(X.row(i))*betat).sum()+omegat(i));
21 }
22 ADREPORT(ldat);
23
24 // ldant - the relative abundance indices for non-target species
25 vector<Type> ldant(n);
26 for(int i=0; i<n; ++i){
27     // Adding omega (random field) and covariates to lambda nt
28     ldant(i)=exp((vector<Type>(X.row(i))*betant).sum()+omegant(i));
29 }
30 ADREPORT(ldant);
31
32 // pnt - the probability of caught non-target species escaping
33 vector<Type> pnt(n);
34 for(int i=0; i<n; ++i){
35     pnt(i)=invlogit(theta);
36 }
37 ADREPORT(pnt);
38
39 DATA_MATRIX(D); // Distance matrix

```

```

40  PARAMETER(lognut);
41  PARAMETER(lognunt);
42  PARAMETER(logPhit);
43  PARAMETER(logPhint);
44  PARAMETER(logSigmat);
45  PARAMETER(logSigmant);
46  // spatial parameters
47  Type nut=exp(lognut); // smoothness parameter
48  Type nunt=exp(lognunt);
49  Type phit=exp(logPhit); // range parameter
50  Type phint=exp(logPhint);
51  Type sigt=exp(logSigmat); // variance
52  Type signt=exp(logSigmant);
53
54  // Covaraince matirx for random field for lambda t
55  matrix<Type> St(n,n);
56  St.setZero();
57  for(int i=0; i<n; ++i){
58      St(i,i) = sigt*sigt;
59  }
60  for(int i=0; i<n; ++i){
61      for(int j=i+1; j<n; ++j){
62          St(i,j) = sigt*sigt*matern(D(i,j), phit, nut); //Matern
63          Covariance Function
64          St(j,i) = St(i,j);
65      }
66  }
67  // Covaraince matirx for random field for lambda nt
68  matrix<Type> Snt(n,n);
69  Snt.setZero();
70  for(int i=0; i<n; ++i){
71      Snt(i,i) = signt*signt;
72  }
73  for(int i=0; i<n; ++i){

```

```

74     for(int j=i+1; j<n; ++j){
75         Snt(i,j) = signt*signt*matern(D(i,j), phint, nunt); //Matern
76         Snt(j,i) = Snt(i,j);
77     }
78 }
79
80 Type a = density::MVNORM(St)(omegat);
81 Type b = density::MVNORM(Snt)(omegant);
82 REPORT(St);
83 REPORT(Snt)
84
85 // Corresponding probabilities for N_b, N_t, N_nt and N_e
86 // N_b is the number of baited hooks
87 // N_t is the number of target species caught
88 // N_nt is the number of non-target species caught
89 // N_e is the number of empty hooks
90 matrix<Type> p(n, n_k);
91 for(int i=0; i<n; i++){
92     p(i,0) = exp(-(ldat(i)+ldant(i))*s(i));
93     p(i,1)=(Type(1)-exp(-(ldat(i)+ldant(i))*s(i)))*(ldat(i)/(ldat(i)
94     +ldant(i)));
95     p(i,2)=(Type(1)-exp(-(ldat(i)+ldant(i))*s(i)))*(ldant(i)/(ldat(i)
96     +ldant(i)))*(Type(1)-pnt(i));
97     p(i,3)=1-p(i,0)-p(i,1)-p(i,2);
98 }
99
100 // The likelihood function for multinomial distribution
101 Type nll = 0;
102 for(int i=0; i < n; i++){
103     vector<Type> p_row=p.row(i);
104     vector<Type> H_row=H.row(i);
105     nll -= dmultinom(H_row,p_row, true);
106 }
107 nll=nll+a+b;

```

```

106 REPORT(a);
107 REPORT(b);
108
109 REPORT(ldat);
110 REPORT(ldant);
111 REPORT(pnt);
112 REPORT(p);
113 REPORT(betat);
114 REPORT(betant);
115 REPORT(omegat);
116 REPORT(omegant);
117 return nll;
118 }

```

- R Code for Data Manipulation:

```

1 ## For 2017
2 # Clear memory
3 rm(list=ls())
4
5 library(TMB)
6 library(fields)
7 library(dplyr)
8 library(ggplot2)
9 library(marmap)
10 library(RandomFields)
11 # Call TMB function value
12 compile("dmulticovariate_Oct.cpp")
13 # Dynamically link the C++ code
14 dyn.load(dynlib("dmulticovariate_Oct"))
15
16 # The halibut data for 2017
17 data_2017 = read.csv("hdata2017.csv", header = T, stringsAsFactors =
    F)
18 head(data_2017)
19 dim(data_2017)
20

```



```
21 st_id = read.csv("HS_STATION_STRATA_2017_18_19.csv", header = F,
  stringsAsFactors = F)
22 names(st_id) = c("YEAR", "STATION", "STRATA")
23 st_id_17 = st_id[which(st_id$YEAR==2017), ]
24 # Merge the data
25 data2017 = merge(data_2017, st_id_17, by = "STATION")
26 head(data_2017)
27 dim(data2017)
28
29 # Drop the stations (523 and 525) with hook = 30
30 not30 = which(data2017$hooks_sampled == 30)
31 data2017[not30,]
32 data2017 = data2017[-c(not30), ]
33 which(data2017$hooks_sampled == 30)
34
35 # Drop the Stratum H13
36 h13_17 = which(data2017$STRATA == "H13")
37 data2017[h13_17,]
38 data2017 = data2017[-c(h13_17), ]
39 which(data2017$STRATA == "H13")
40 dim(data2017)
41
42 # Read in the block id for the prediction
43 blockid = read.csv("blockIDkey.csv", header = T)
44 blockid[1:4, ]
45 dim(blockid)
46 # Drop the Stratum H13
47 h13_id = which(blockid$s.id == 1.3)
48 blockid[h13_id,]
49 blockid = blockid[-c(h13_id), ]
50 which(blockid$s.id == 1.3)
51 dim(blockid)
52
53 library(sp)
54 library(rgdal)
```

```

55 library(INLA)
56 #To switch for lat-long and project it on a flat surface since the
    earth is a global
57 prj4s=CRS("+init=epsg:4326")
58 utm.prj4s=CRS("+init=epsg:32619")
59 loc_2017=cbind(data2017$P1LONG,data2017$P1LAT)
60 loc_17=cbind(data2017$P1LONG,data2017$P1LAT)
61 loc_17=SpatialPoints(loc_17,proj4string = prj4s)
62 loc_17=spTransform(loc_17,utm.prj4s)
63 data2017$P1LONG_proj=loc_17@coords[,1]
64 data2017$P1LAT_proj=loc_17@coords[,2]
65
66 # Standardize loc for choosing a starting value for phi
67 spool_17 = sqrt(((length(loc_2017[,1])-1)*var(loc_17@coords[,1])+
    length(loc_2017[,2])-1)*var(loc_17@coords[,2]))/(length(loc_
    2017[,1])+length(loc_2017[,2]))
68 data2017$loc1=(loc_17@coords[,1]-median(loc_17@coords[,1]))/spool_17
69 data2017$loc2=(loc_17@coords[,2]-median(loc_17@coords[,2]))/spool_17
70 dismat_17=cbind(data2017$loc1,data2017$loc2)
71 # Distance matrix
72 Dist_17= as.matrix(dist(dismat_17))
73
74 # Put the required data into a matrix
75 data2017$empty=data2017$empty_unbaited+data2017$broken_hook
76 H_17=with(data2017, cbind(empty_baited, halibut, other_species,
    empty))
77 # Soak time
78 s_17=data2017$SOAKMINP3P1
79 s_17=as.vector(s_17)
80
81 # Logit function
82 logitp=function(p){log(p/(1-p))}
83 # Inverse logist function
84 logitpi=function(t){exp(t)/(1+exp(t))}
85

```

```

86 # The intercept (column of ones)
87 colone_17=rep(1,nrow(data2017))
88 X_170=as.matrix(colone_17)
89 data_170=list(H=H_17,s=s_17,X=X_170,D=Dist_17)
90
91 param_17tnt0 = list()
92 # Initial values for lambda.t, lambda.nt and pnt
93 # Use the estimated values as the starting points
94 param_17tnt0$betat = log(1.878483e-05)
95 param_17tnt0$betant = log(0.001941084)
96 param_17tnt0$theta = logitp(0.8795768)
97 # Random field
98 param_17tnt0$omegat = rep(0,nrow(H_17))
99 param_17tnt0$omegant = rep(0,nrow(H_17))
100 # Smoothness parameter
101 param_17tnt0$lognut = 0
102 param_17tnt0$lognunt = 0
103 # Range parameter
104 param_17tnt0$logPhit = log(diff(range(dismat_17[,1]))*diff(range(
    dismat_17[,2])))
105 param_17tnt0$logPhint = log(diff(range(dismat_17[,1]))*diff(range(
    dismat_17[,2])))
106 # Variance
107 param_17tnt0$logSigmat = 0
108 param_17tnt0$logSigmant = 0
109
110 newlist_170=list(lognut=factor(NA),lognunt=factor(NA))
111 # Construct an R object (f) that represents our C++ function
112 # use map and newlist debug
113 # Fix lognu at 0
114 f_17tnt0 = MakeADFun(data_170,param_17tnt0,random=c("omegat",
    omegant"), DLL="dmulticovariate_Oct",map=newlist_170,silent=TRUE)
115 fit_17tnt0 = nlminb(f_17tnt0$par,f_17tnt0$fn,f_17tnt0$gr)
116 # Calculate standard deviations of all model parameters
117 sdr_17tnt0 = sdreport(f_17tnt0)

```

```

118 # The estimated parameters and corresponding standard deviations
119 sum_sdr170 = summary(sdr_17tnt0)
120
121 # C++ file which defines the objective function (usually the
      negative log likelihood)
122 # R file which sets up data, calls the C++ function, and minimizes
      the objective function.
123 # Minimized negative log likelihood
124 fit_17tnt0$objective
125
126 # Estimated Relative Abundance Indices
127 # 2017 without covariate
128 # For target species
129 head(sum_sdr170[row.names(sum_sdr170) %in% "ldat", ])
130 dim(sum_sdr170[row.names(sum_sdr170) %in% "ldat", ])
131 lamt_est_170 = data.frame(sum_sdr170[row.names(sum_sdr170) %in% "
      ldat", ])
132 head(lamt_est_170)
133 data2017$lamt_se_170 = lamt_est_170$Std..Error
134 data2017$lamt_170 = lamt_est_170$Estimate
135 # For non-target species
136 head(sum_sdr170[row.names(sum_sdr170) %in% "ldant", ])
137 dim(sum_sdr170[row.names(sum_sdr170) %in% "ldant", ])
138 lamnt_est_170 = data.frame(sum_sdr170[row.names(sum_sdr170) %in% "
      ldant", ])
139 head(lamnt_est_170)
140 data2017$lamnt_se_170 = lamnt_est_170$Std..Error
141 data2017$lamnt_170 = lamnt_est_170$Estimate
142
143 ### Plot the Relative Abundance Indices
144 n_17 = c(1:nrow(data2017))
145 indices_17t = data2017$lamt_170
146 se_17t = data2017$lamt_se_170
147
148 # For target species

```

```

149 data_17t = data.frame(n=n_17, indices_t=indices_17t, se=se_17t)
150 p_17t = ggplot(data_17t, aes(x=n, y=indices_t)) + geom_point() +
      geom_errorbar(aes(ymin=indices_t-se, ymax=indices_t+se), width
      =0.5) + labs(x = "Survey Station", y = "Relative Abundance
      Indices for Target Species")
151 p_17t
152
153 # For non-target species
154 indices_17nt = data2017$lamnt_170
155 se_17nt = data2017$lamnt_se_170
156 data_17nt = data.frame(n=n_17, indices_nt=indices_17nt, se2=se_17nt)
157 p_17nt = ggplot(data_17nt, aes(x=n, y=indices_nt)) + geom_point()+
      geom_errorbar(aes(ymin=indices_nt-se2, ymax=indices_nt+se2),
      width=0.5) + labs(x = "Survey Station", y = "Relative Abundance
      Indices for Non-target Species")
158 p_17nt

```

A.4 Interpolation of Estimated Relative Abundance Indices and Random Field

The following codes are used for interpolating the random field and relative abundance indices for target species over the entire block id without stratum H13.

```

1 # Continuing from the code of MEM1 section
2 ## For 2017
3 ##### With R function
4 # For lambda t
5 # Range parameter for lambda t
6 logphit_170 = sum_sdr170[row.names(sum_sdr170) %in% "logPhit", ][1]
7 logphit_170
8 phit_170 = exp(as.numeric(logphit_170))
9 # Variance for lambda t
10 logsigmat_170 = sum_sdr170[row.names(sum_sdr170) %in% "logSigmat",
      ][1]
11 logsigmat_170
12 var_170 = (exp(as.numeric(logsigmat_170)))^2

```

```

13 # Random field for lambda t
14 omt_170 = data.frame(sum_sdr170[row.names(sum_sdr170) %in% "omegat",
15 ])
16
17 #To switch for lat-long and project it on a flat surface since the
18 #earth is a global
19 loc_id=cbind(blockid$lon.DecDeg, blockid$lat.DecDeg)
20 loc_idtrans=NULL
21 loc_idtrans=SpatialPoints(loc_id,proj4string = prj4s)
22
23 # Standardize block id
24 loc_idtransnorm_17=NULL
25 loc_idtransnorm_17$lon=(loc_idtrans@coords[,1]-median(loc_17@coords
26 [,1]))/spool_17
27 loc_idtransnorm_17$lat=(loc_idtrans@coords[,2]-median(loc_17@coords
28 [,2]))/spool_17
29 mloc_idtransnorm_17=as.matrix(cbind(loc_idtransnorm_17$lon,loc_
30 idtransnorm_17$lat))
31
32 # Whittle-Matern Covariance Model
33 modelt_170 = RMwhittle(nu=1,scale=phit_170,var=vart_170)
34 # Interpolate omega (random field)
35 predt_170 = RFinterpolate(modelt_170,x=mloc_idtransnorm_17,
36 data=data.frame(x=dismat_17,data=est_omt_170),spConform=
37 FALSE)
38 lambdat_170=exp(as.numeric(sum_sdr170[row.names(sum_sdr170) %in% "
39 betat", ][1])+predt_170)
40 blockid$intmega_17=predt_170
41 blockid$intltdat_17=lambdat_170
42
43 # Plot the interpolated random field
44 col0=colspace::diverge_hsv(256)
45 col1=c(col0[c(seq(1,127,by=3),128,129:256)])

```

```

41 # Random field
42 quilt.plot(loc_id, matrix(predt_170), xlab="long", ylab="lati", main=
    expression(paste("Heat map of random field for ", lambda, "_t for
    2017")), nx=220, ny=220, col=col2, zlim=c(-1.4, 4.2))
43 world(add=TRUE, col="black", lwd=1)
44 points(loc_2017, type="p", pch=20, cex=0.6)
45
46 # Plot the interpolated relative abundance indices
47 col3=c(col10[c(128, 129:256)])
48 quilt.plot(loc_id, matrix(lambdat_170), main=expression(paste("
    Estimated ", lambda, "_t for 2017")), xlab="long", ylab="lati", col=
    col3, zlim=c(0, max(lambdat_170)))
49 world(add=TRUE, col="black", lwd=1)
50 points(loc_2017, type="p", pch=20, cex=0.6)
51
52 ##### With Theoretical Calculation
53 # Covariance matrix for Whittle-Matern model with 2017 survey
    location
54 covy=RFcovmatrix(modelt_170, x=dismat_17)
55 # Covariance matrix for Whittle-Matern model with block id location
56 covy0=RFcovmatrix(modelt_170, x=mloc_idtransnorm_17)
57 # Distance between block id location and 2017 survey location
58 disy0y=rdist(mloc_idtransnorm_17, dismat_17)
59 # Computes the empirical cross-covariance function for given spatial
    data.
60 # Covariance for Whittle-Matern model with block id location
61 # and Whittle-Matern model with 2017 survey location
62 covy0y=RFcov(modelt_170, dist=as.vector(disy0y), dim=2)
63 covy0y=matrix(covy0y, ncol=nrow(covy))
64 # Interpolate omega (random field)
65 y0=covy0y%*%solve(covy, est_omt_170)
66 # Check R function result and theoretical calculation results
67 plot(y0, predt_170)
68 sum((y0-predt_170)^2)

```

A.5 Aggregation with Dirichlet Method

The following codes are used for calculating weighted average of relative abundance indices for halibut using Dirichlet method without stratum H13.

```

1 # Continuing from the code of MEM1 section
2 ## For 2017
3 ## Dirichlet
4 library(gissr)
5 library(deldir)
6 library(spatstat)
7 library(alphahull)
8
9 # Read in block id
10 bid = read.csv("blockIDkey.csv", header = T)
11 loc_id = cbind(bid$lon.DecDeg, bid$lat.DecDeg)
12 # Plot block id
13 plot(loc_id, xlab="Long", ylab="Lat", main="blockIDkey", pch=20)
14
15 # With Alpha Hull
16 # Find the boundary points using ashape
17 # Alpha controls the detail of the boundary
18 bound_a = ashape(loc_id, alpha = 0.084)
19 bound_a_index = bound_a$alpha.extremes
20 plot(loc_id, xlab="Long", ylab="Lat", main="blockIDkey and boundary",
21       , pch=20)
22
23 ### alpha = 0.084
24 # Boundary points and survey stations
25 bound_a_pos = loc_id[bound_a_index, ]
26 plot(bound_a_pos, pch=20)
27 # Extracted block id boundary and imported boundary
28 plot(bound_a_pos, xlab="Long", ylab="Lat", main=" Survey Area
29       Boundary Coordinate", pch=20)
30 points(surveyPolyLL$X, surveyPolyLL$Y, col="red", pch=20)

```



```

30
31 # To find the corresponding index
32 bound_a_pos = data.frame(long=bound_a_pos[,1], lat=bound_a_pos[,2])
33 bound_a_pos$num = 1:length(bound_a_pos$long)
34
35 # Fix the boundary
36 dim(bound_a_pos)
37 bound_a_pos = bound_a_pos[-c(12,15:21,57,58,579,581,588,594,
38 595,598,599,601,604,605,668,669,828,834,840,844,845,841,835,
39 773,751,755,320,311:313,326,340,357,361,364,365,360,356,353,
40 336,325,452,472,495,507,678,685,686,634,630,635,639,638,640,
41 323,316,306),]
42 dim(bound_a_pos)
43 plot(bound_a_pos$long, bound_a_pos$lat, pch=20)
44
45 # To switch for lat-long for block id
46 # and project it on a flat surface since the earth is a global
47 tf_bound_a_pos = SpatialPoints(bound_a_pos,proj4string = prj4s)
48 tf_bound_a_pos = spTransform(tf_bound_a_pos,utm.prj4s)
49 tf_bound_a_pos = data.frame(long = tf_bound_a_pos@coords[,1], lat =
50   tf_bound_a_pos@coords[,2])
51 plot(tf_bound_a_pos)
52
53 # List the boundary points in anit-clockwise order
54 tf_bound_a_pos2 = sort_points(tf_bound_a_pos, "lat", "long",
55   clockwise = FALSE)
56 tf_bound_a_pos2 = data.frame(long = tf_bound_a_pos2$long, lat = tf_
57   bound_a_pos2$lat, num=1:length(bound_a_pos$long))
58
59 # Creates an object of class "ppp" representing a point pattern
60   dataset in the two-dimensional plane
61 pp_17=ppp(data2017$P1LONG_proj, data2017$P1LAT_proj, window=owin(
62   poly=list(x=tf_bound_a_pos2$long, y=tf_bound_a_pos2$lat)))
63 dpp_17 = dirichlet(pp_17)

```

```

60 ## Dirichlet plot
61 tf_bound_a_pos22 = tf_bound_a_pos2[926, ]
62 tf_bound_a_pos22[1:64, ] =
63   tf_bound_a_pos2[c(918,921,925,929,932,936,935,934,933,931,
64   930,927,926,924,923,922,920,919,928,937:953,955,954,956,
65   958,960,957,959,40,48,55,61,59,66,67,76,84,87,94,95,91,
66   97,99,101,102,105,106,110,124), ]
67 tf_bound_a_pos22[65:160, ] =
68   tf_bound_a_pos2[c(127,117,120,123,113,116,118,122,125,115,
69   129,134,138,142,148,144,140,136,131,111,119,128,141,146,
70   150,152,156,158,162,163,166,167,170,176,179,181,185,186,
71   189,190,194,195,191,196,199,201,202,205,215,214,213,212,
72   207,211,210,209,206,204,198,192,187,182,177,174,171,164,
73   161,154,151,145,135,133,130,112,108,107,104,96,92,90,89,
74   83,80,77,75,70,68,64,62,60,57,54,52,46,43,37), ]
75 tf_bound_a_pos22[161:250, ] =
76   tf_bound_a_pos2[c(36,27,29,28,23,17,14,12,10,13,15,19,22,
77   20,18,6,5,2,963,961,962,1,3,4,7:9,11,16,21,24,25,26,33,35,
78   32,30,31,34,38,39,42,45,49,41,44,47,50,51,53,56,58,63,65,
79   69,73,81,86,78,71,72,74,79,82,85,88,93,98,100,103,109,114,
80   121,126,132,137,139,143,147,149,153,155,157,159,160,165,
81   168,169,172,173), ]
82 tf_bound_a_pos22[251:371, ] =
83   tf_bound_a_pos2[c(175,178,180,183,184,188,193,197,200,203,
84   208,216:235,250,260,265,272,280,283,294,295,292,290,286,
85   285,288,287,281,278,277,276,274,271,270,269,266,261,254,
86   247,245,240,236,241,237,242,238,243,248,244,239,246,251,
87   255,252,256,253,249,257:259,262:264,267,268,273,275,279,
88   282,284,289,291,293,296:306,309,307,308,310:321,323,328,
89   332,331), ]
90 tf_bound_a_pos22[372:444, ] =
91   tf_bound_a_pos2[c
92     (327,325,322,324,330,326,329,333:349,352,351,350,353:398), ]
93 tf_bound_a_pos22[445:545, ] =
94   tf_bound_a_pos2[c(401,400,399,402:405,426,447,460,469,486,

```

```

94 491,512,517,522,530,523,531,524,519,516,526,535,540,549,
95 557,559,555,563,571,576,577,574,560,562,570,572,567,564,
96 566,569,558,551,548,543,545,542,544,541,538,537,532,527,
97 533,528,525,521,518,508,514,510,506,511,507,504,499,505,
98 501,496,502,497,503,498,493,487,483,480,484,482,479,477,
99 476,474,473,470,466,461,458,457,463,459,455,453,445,444,
100 449,446,451,450,441), ]
101 tf_bound_a_pos22[546:615, ] =
102   tf_bound_a_pos2[c(439,437,443,440,438,436,427,409,406,407,
103   413,431,435,434,424,420,428,425,422,418,415,411,408,417,
104   414,410,421,416,412,423,419,432,430,429,433,442,448,452,
105   454,456,462,464,465,467,468,471,472,475,478,481,485,488,
106   489,490,492,494,495,500,509,513,515,520,539,534,536,539,
107   546,547,550,552), ]
108 tf_bound_a_pos22[616:688, ] =
109   tf_bound_a_pos2[c(553,554,556,561,565,568,573,575,578:612,
110   620,619,618,617,616,615,614,613,627,626,625,624,623,622,
111   621,633,632,631,630,629,628,639,638,637,636,635,634,645,
112   644,643), ]
113 tf_bound_a_pos22[689:785, ] =
114   tf_bound_a_pos2[c(642,641,640,650,649,648,647,646,656,655,
115   654,653,652,651,657:739), ]
116 tf_bound_a_pos22[786:874, ] =
117   tf_bound_a_pos2[c(750,756,764,769,780,777,776,773,770,779,
118   766,765,761,754,752,749,746,744,751,757,760,768,778,775,
119   772,767,763,759,758,747,745,742,741,740,743,753,748,755,
120   762,774,771,781,783,782,784:786,790,803,815:818,811,810,
121   806,808,798,800,802,792,794:796,799,804,805,807,797,791,
122   787:789,793,801,809,812,819,822,824,825,823,814,813,821,
123   820,826,828,827), ]
124 tf_bound_a_pos22[875:963, ] = tf_bound_a_pos2[c
      (829:913,915,914,916:917), ]
125
126 tf_bound_a_pos22 = data.frame(long = tf_bound_a_pos22$long, lat = tf
      _bound_a_pos22$lat,num=1:length(tf_bound_a_pos22$long))

```

```

127
128 # Creates an object of class "ppp" representing a point pattern
      dataset in the two-dimensional plane
129 pp_17=ppp(data2017$P1LONG_proj, data2017$P1LAT_proj, window=owin(
      poly=list(x=tf_bound_a_pos22$long, y=tf_bound_a_pos22$lat)))
130 dpp_17 = dirichlet(pp_17)
131 plot(dpp_17,border="black",main="")
132 plot(pp_17,add=TRUE,chars=20,col="red", main="Dirichlet Tiles Plot
      for 2017")
133
134 ## Area
135 wts_17 = sapply(tiles(dpp_17),area.owin)
136 cat("Sum of weights:\n")
137 print(sum(wts_17)/1000000)
138 # Target
139 dweight_ave_17t = sum((data2017$lamt_170)*wts_17)/sum(wts_17)
140 dweight_ave_17t
141 # Standard error calculation
142 wts_17 = as.matrix(wts_17)
143 # wi/sum(wi)
144 wts_172 = wts_17/(sum(wts_17))
145 # covariance matrix for estimated lambda t
146 cov_17 = sdr_17tnt0$cov
147 cov_17t = cov_17[1:nrow(data2017), 1:nrow(data2017)]
148 # plot(((data2017$lamt_se_170)^2,diag(cov_17t))
149 sum(((data2017$lamt_se_170)^2-diag(cov_17t)))
150 # Standard error for Dirichlet method
151 se2_dir_17t = t(wts_172)%*%cov_17t%*%wts_172
152 sqrt(se2_dir_17t)

```