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

Canadian Science Advisory Secretariat (CSAS)

Research Document 2026/071

Maritimes Region

Estimating Atlantic Cod Bycatch in Commercial Lobster Fisheries in NAFO Divisions 4X5Y

[1]Yihao Yin, [2]Irene Andrushchenko, and [1]Adam M. Cook

[1]Fisheries and Oceans Canada
Bedford Institute of Oceanography
P.O. Box 1006, 1 Challenger Drive
Dartmouth, Nova Scotia, B2Y 4A2

[2]Fisheries and Oceans Canada
Saint Andrews Biological Station
125 Marine Science Drive
St. Andrews, New Brunswick, E5B 0E4

October 2026

Canada

Foreword

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

Published by:

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

http://www.dfo-mpo.gc.ca/csas-sccs/
DFO.CSAS-SCAS.MPO@dfo-mpo.gc.ca



© His Majesty the King in Right of Canada, as represented by the Minister of the Department of Fisheries and Oceans, 2026

This report is published under the Open Government Licence - Canada

ISSN 1919-5044
ISBN 978-1-100-00649-9 Cat. No. Fs70-5/2026-071E-PDF

Correct citation for this publication:

Yin, Y., Andrushchenko, I., and Cook, A.M. 2026. Estimating Atlantic Cod Bycatch in Commercial Lobster Fisheries in NAFO Divisions 4X5Y. DFO Can. Sci. Advis. Sec. Res. Doc. 2026/071. iv + 69 p.

Aussi disponible en français :

Yin, Y., Andrushchenko, I., et Cook, A.M. 2026. Estimation des prises accessoires de morue franche dans la pêche commerciale du homard des divisions 4X5Y de l'OPANO. Secr. can. des avis sci. du MPO. Doc. de rech. 2026/071. iv + 70 p.

TABLE OF CONTENTS

ABSTRACT iv
1. INTRODUCTION 1
2. DATA 1
2.1. DATA SOURCES AND DESCRIPTION 1
2.1.1. Summer Ecosystem Research Vessel Survey (RV) 1
2.1.2. Inshore Lobster Trawl survey (ILTS) 2
2.1.3. At-Sea Sampling Programs (AS) 2
2.1.4. Commercial Lobster Fishing Effort (CLFE) 3
2.1.5. North Atlantic Ocean Bathymetry 4
2.2. GENERATING PREDICTORS FOR BYCATCH ANALYSIS 4
2.2.1. Depth 4
2.2.2. Cod Density 4
2.3. SUMMARY OF PREDICTORS 5
3. METHOD 6
3.1. EXPLORATORY ANALYSIS 6
3.2. MEAN ESTIMATOR (ME) 6
3.3. NON-PARAMETRIC MODEL 7
3.4. GENERALIZED LINEAR MODELS 8
3.5. SPATIAL MODEL 8
3.6. MODEL SELECTION 9
3.7. PREDICTING TOTAL BYCATCH 10
4. RESULTS 11
4.1. MODEL SELECTION RESULTS 11
4.2. PREDICTED TOTAL BYCATCH 11
4.3. SOURCES OF UNCERTAINTY 12
REFERENCES CITED 12
FIGURES 15
TABLES 60

ABSTRACT

The stock status update for Atlantic Cod in Northwest Atlantic Fisheries Organization (NAFO) Divisions 4X5Y (southern Scotian Shelf and Bay of Fundy, Figure 1) for 2022 identified high natural mortality as the cause of persistent low productivity for the stock including mortality due to incidental catches (bycatch) from all non-groundfish fisheries operating within the region. In compliance with DFO's *Policy on Managing Bycatch*, a methodology for quantifying the interaction of Cod with the Lobster fishery within 4X5Y is proposed. The approach utilized and integrated data from multiple sources including the Maritimes Ecosystem Research Vessel (RV) Survey, Inshore Lobster Trawl Survey (ILTS), at-sea sampling of commercial Lobster fishing activities, and reports of commercial Lobster fishing effort. The analysis takes into account the non-uniform distribution of both Cod biomass and Lobster fishing effort across the 4X5Y area with considerations of related environmental processes. Vigorous validation and testing were conducted to evaluate performance of the proposed approach on the predictive ability on total Cod bycatch in the commercial Lobster fishery as well as its possible limitations. Subsequent work focusing on the rate of mortality applied to the bycatch estimates and incorporation into the 4X5Y Cod assessment will be presented at a future peer review meeting.

1. INTRODUCTION

The stock status update for Atlantic Cod in Northwest Atlantic Fisheries Organization [NAFO] Divisions 4X5Y (southern Scotian Shelf and Bay of Fundy, Figure 1) for 2022 identified high natural mortality as the cause of persistent low productivity for the stock (DFO 2023a). The fishery removals in the current Cod assessment model include all known removals by the groundfish fleet in 4X5Y, with all other unaccounted-for mortality in NAFO Divisions 4X5Y categorized as natural mortality in the model. Therefore, the current high estimate of natural mortality for 4X5Y Cod includes, among other things, mortality due to incidental catches of Atlantic Cod from all non-groundfish fisheries operating within NAFO Divisions 4X5Y (DFO 2023a).

Fishing gear is designed to capture specific target species, while simultaneously minimizing interaction with non-target species. In cases where this interaction is unavoidable, the unintentional catch of a species by a fishing gear is referred to as “bycatch”. The management of bycatch in Canadian fisheries falls under the *Policy on Managing Bycatch*, which aims to account for all retained and non-retained bycatch in Canadian fisheries (DFO 2013). As a demersal species, Cod is expected to interact with all fisheries targeting the benthic or near benthic habitats within NAFO Divisions 4X5Y, including groundfish, scallop and Lobster.

A previous peer review of methodologies to estimate the interaction of Cod with the Lobster fishery was conducted in 2022, but was limited to Lobster Fishing Areas (LFAs) 27, 31A, 31B, 33, 34 and 35 (Figure 1; DFO 2023b). The 4X5Y Cod assessment unit encompasses LFA 34, 35, and a large portion of 33, but also includes areas 36, 37, 38, and 41 (Figure 1). The spatial misalignment between the two necessitated the re-analysis of the Cod bycatch estimates from the previous peer review in order to align with the 4X5Y assessment unit to facilitate its incorporation into the Cod assessment (see Figure 2 for the overlapping area between NAFO Divisions 4X5Y and LFAs 33-41).

In compliance with DFO’s *Policy on Managing Bycatch*, a methodology for quantifying the interaction of Atlantic Cod with the Lobster fishery within NAFO Divisions 4X5Y is proposed. The approach takes into account the non-uniform distribution of both Cod abundance and Lobster fishing effort across the 4X5Y area, utilizing data from multiple sources including the Maritimes Summer Ecosystem Research Vessel Survey (herein referred to as the RV Survey), Inshore Lobster Trawl Survey (ILTS), at-sea sampling of commercial Lobster fishing activities, and reports of commercial Lobster fishing effort. Subsequent work focusing on the rate of mortality applied to the bycatch estimates and incorporation into the 4X5Y Cod assessment will be presented at a future peer review meeting.

2. DATA

2.1. DATA SOURCES AND DESCRIPTION

2.1.1. Summer Ecosystem Research Vessel Survey (RV)

Fisheries and Oceans Canada (DFO) has conducted the RV Survey in the Maritimes Region using a standardized protocol since 1970 (DFO 2024) and these data can provide valuable insight into both Cod abundance and distribution within NAFO Divisions 4X5Y. The RV survey follows a stratified random survey design (Figure 3). Summer missions typically occur during the summer months with 4X5Y being sampled in July. The survey uses a bottom trawl gear fishing at depths between 30 and 400 m, recording the total weight of fish and invertebrates caught per set, as well as various biological and environmental variables (DFO 2024). Sampling of Cod

includes individual fish lengths, weights, maturities, and ages recorded from each set at a one per 3 cm bin aggregation. In times of vessel and gear changes, comparative fishing experiments and analysis were conducted to maintain consistency (Yin et al. 2025). The RV survey data from NAFO Divisions 4X5Y from 2016 to 2024 were used in this study to estimate a relative Cod biomass density distribution within the area. Although the RV Survey time series is available back to the 1970s, the current work focused on 2016–2024 due to the availability of the at-sea sampling dataset (Section 2.1.3). A summary of the RV Survey data is shown in Table 1.

2.1.2. Inshore Lobster Trawl survey (ILTS)

The ILTS evolved from the Individual Transferable Quota (ITQ) Survey in 2013. The ITQ Survey began in 1995 and was a fixed station mobile gear survey to provide abundance indices for groundfish stock assessments in NAFO Division 4X (Denton, 2020). The ILTS takes place during June and July and covered an average of 181 stations per year in recent years (Figures 4 and 5). Similar to the RV Survey, comparative fishing experiments and analysis were conducted for gear changes which occurred in 2016 (Cook et al. in review). The survey samples Cod and characterizes catch numbers and length frequency (Denton, 2020) that can be used to derive catch weights. Both the RV Survey and the ILTS have been using the same bottom trawl in recent years. The ILTS survey complements the RV Survey as it samples a larger portion of the near shore habitat as well as having spatial overlap in the mid-shore and offshore areas. This sampling in nearshore habitat is important for characterizing the Cod distribution in areas where the Lobster fishery occurs (see Figure 5 for the annual overlap between the two surveys with random stations of RV Survey data and fixed stations of ILTS data). A summary of the ILTS Cod catch data are presented in Table 2.

2.1.3. At-Sea Sampling Programs (AS)

Bycatch in Canadian fisheries is often not well documented and is not collected uniformly across regions or for various target species. The *Policy on Managing Bycatch* (DFO 2013) identified a need to systematically account for bycatch across various fisheries, and the implementation of objectives and measures where bycatch is a conservation concern. The Maritimes Inshore Lobster Fishery is a priority for this analysis because of the risks the fishery presents to species retained for use as bait (e.g., Jonah Crab) and the bycatch of depleted groundfish species (e.g., Atlantic Cod and Cusk). In particular, the 4X5Y Atlantic Cod Rebuilding Plan committed fisheries management and Science to collecting bycatch data in the Lobster fisheries in LFAs 33–38 (DFO 2021). In addition to the policy mandate, implementing a system to collect bycatch data were also a condition of the Inshore Lobster Fishery's Marine Stewardship Council certification (Criquet et al. 2015).

In 2014, DFO published an overview of bycatch and discards from LFAs 27–34 with a primary focus on obtaining information on incidental catch, particularly species listed under the *Species at Risk Act* (SARA) and the Committee on the Status of Endangered Wildlife in Canada (COSEWIC) that were captured in commercial Lobster traps (Pezzack et al. 2014). Cod was included in this overview. The study used at-sea observer catch data on these species to estimate the level of discard, by weight, with the majority of the samples collected in LFAs 33 and 34.

In 2018, a standardized at-sea data collection program was introduced in LFAs 33, 34, and 35. The program provided fishers the option to work with at-sea observer companies or an association (Southwest Lobster Science Society; SWLSS) that provided technicians to collect bycatch information during commercial fishing operations in the 2019 fishing season. A preliminary target to collect data from 1% of commercial fishing trips within each LFA in the Inshore Lobster Fishery.

The study documented herein uses at-sea sampling (AS) data from 2016–2024 for analysis and estimation of total Cod bycatch. AS coverage was very limited prior to the pilot program (mostly in LFA 41 before 2018). Data collected in 2024 are not complete as data after October is still undergoing quality control. Figures 6 and 7 demonstrate the annual spatial and temporal coverages of the AS data. Estimates for years prior to 2019 largely depend on extrapolation assuming stationarity of the underlying data-generating process. Due to large uncertainty in these estimates, interpretation and utilization of these estimates should proceed with caution. Estimates for 2024, as noted above, are based on a partial data set for the year, and therefore should be updated upon completion of the quality control of the 2024 data. Additionally, sampling bias (AS data as a subset of the entire commercial Lobster fishery and thus representative of the latter) could introduce inherent bias in estimation of total Cod bycatch. For example, DFO (2023b) assessed AS coverage relative to the fishery for many LFAs during the pilot program and found some LFAs feature small bias in representativity in terms of Lobster catches. Some types of bias could be alleviated via model approaches (e.g., a spatial model to address spatial autocorrelation between close locations), but others may be difficult to evaluate (e.g., possible behavioral bias when being observed). Overall, Cod bycatch patterns feature small annual variations (Figure 8), and a summary of data indicates relatively consistent trends where proportions of zero Cod bycatch are around 98% and average positive bycatch amounts are around 1 kg per trap haul (Figure 9), despite sample size variations across the years.

A summary of AS data used in this study is in Table 3. A total of 113,291 traps were sampled during 2016–2024, with Cod observed in 2,430 traps, meaning nearly 98% of traps did not have Cod inside the trap. However, in some cases Cod may have entered the trap and were completely consumed by other animals. More detailed information regarding sampling of species inside the traps and animal conditions, etc., can be found in Denton (2020). The individual trap-level data for Cod bycatch exhibit high variations with an average coefficient of variation (CV) of 8.7, suggesting an extremely long-tailed distribution of the quantity of interest, i.e., Cod bycatch.

2.1.4. Commercial Lobster Fishing Effort (CLFE)

The fishery level estimates of Cod bycatch depend on, and are scaled up by, the total commercial Lobster fishing efforts (CLFE). CLFE data rely on the mandatory logbook data and sales slips provided from commercial harvesters. These two data sources represent the most comprehensive source of information on the Lobster fisheries in the Maritimes Region. Details of these data sources can be found in Cook et al. (2024). The CLFE data used in this study was provided by Cook et al. (2024) following quality control, analysis, and validation processes.

For inshore Lobster areas considered in this study (LFAs 33–38, Figure 1), fishing effort was reported in aggregation both spatially and temporally. Spatial aggregation is based on a set of grids (i.e., Lobster grids, Figure 10) and temporal aggregation is based on the week of Lobster fishing season which varies according to the LFA; this gives an estimated total number of traps for each Lobster grid during each week of the season. To reconcile this mis-alignment between resolutions of CLFE data and modelled AS data, we developed the bycatch analysis based on the original higher resolution of data, i.e., given exact locations and dates, and subsequently predicted to the lower resolution data with averaging. This could add uncertainty depending on distribution of effort within the Lobster grid and should be assessed in future studies. For offshore areas (i.e., LFA 41), both AS and CLFE contain high resolution data so there is no mis-alignment for model-based approaches. Notably all effort in LFA 41 is from a single license holder, so figures and numbers pertaining to this area were either redacted or removed in this document to comply with confidentiality requirements.

Overall, spatial and temporal patterns of fishing effort are relatively consistent during the study period and throughout the area, providing some evidence to support the stationarity assumption (Figures 11 and 12; note that data for 2024 are partial where a majority of trips after October were undergoing quality control at the time of assessment). Table 4 summarizes the effort data aggregated to various area specifications. The AS coverage rate is well below the 1% target of the program (Tables 3 and 4).

2.1.5. North Atlantic Ocean Bathymetry

Model bathymetry taken from the ORCA12 model of Mercator-Ocean, France (Treguier et al, 2017) was processed by Wang et al. (2018) where partial cell scheme was applied to better represent the bathymetry. The bathymetry data for NAFO Divisions 4X5Y were incorporated in both the model that estimates Cod population density distribution and the model that estimates Cod bycatch in the Lobster commercial fishery in this study.

2.2. GENERATING PREDICTORS FOR BYCATCH ANALYSIS

Several relevant predictors were used for estimating Cod bycatch. Data for these predictors were either directly observed or generated via separate modeling approaches.

2.2.1. Depth

Depth is an informative predictor for many ecological processes and was collected during most sampling programs, including both the RV Survey and the ILTS. However, these sources often have missing data and the collected data were often subject to rough bottoms and incidental observation errors. Additionally, this study integrated various data sources via multiple stages of statistical modeling where depth was utilized. To ensure consistency and accuracy of the depth data used in modeling, the high-resolution North Atlantic ocean bathymetry data were used from Wang et al. (2018). The resolution and gridding scheme of the bathymetry, i.e., the 1/12 degree ORCA12 model grid, was also used for modeling and result presentation in this study for both consistency and readability reasons (hereafter referred to as “bathymetry grid” in order to differentiate from other modeling grids mentioned in this study). Figure 13 presents these depth data and Figure 14 shows the difference between these data and observed depths in the ILTS.

2.2.2. Cod Density

In order to generate a predictor that represents relative Cod population density levels across both space and time, and that is useful for the subsequent Cod bycatch analysis, a spatial model was developed to integrate the RV Survey and the ILTS. Data from each survey were first analyzed separately to incorporate comparative fishing (Yin et al. 2025) and conversion from catch numbers to weights (Figures 15 and 16 show Cod catch weights per tow for the two surveys, respectively). The length-weight relationship used for conversion was from Andrushchenko et al. (2022) and is dependent on the season (see Table 5 for details). In addition, weights for damaged Cod from the ILTS were imputed assuming a blanket 30% biomass loss (i.e., weight was scaled up by a factor of 1.3 post length-weight conversion for a damaged fish). Catches from the two surveys indicated different trends, possibly due to differences in spatial coverages that correlate to different Cod distributions (Figure 17).

To develop a spatial biomass density distribution model for 4X5Y Cod, a modeling area needed to be defined first. A polygon was derived that covers all the AS locations within 4X5Y and LFAs 33–41, with a buffer zone of 2 km and simplified to reduce edge effects (Figure 18). This boundary was necessary in this study to avoid misrepresentation of spatial correlation between locations that are separated by land masses. The RV Survey locations (random survey design,

therefore locations vary every year), the set of fixed survey locations for ILTS (same every year), and a thinned dataset containing 1,000 AS locations (i.e., randomly subsampled from all locations in the AS dataset to reduce computation while maintaining representation of its spatial coverage, as well as patterns) were combined. A distance-based K-means clustering algorithm was used to aggregate and associate the locations to 60 “knots” (see definition for knots in Lindgren and Rue 2015). The spatial model relies on these knots, similar to strata, and builds on spatial and spatiotemporal autocorrelation upon them (Figure 18).

The following spatial model was preferred among several tested models with various settings pertaining to different underlying assumptions of spatial and spatiotemporal variations and covariates (models and model selection based on statistical evidence are listed in Table 6):

$$log(Density) = \alpha(\text{Survey}) + \alpha(Year) + \text{s}(\text{Depth}) + \beta(\text{S}) + \beta(\text{ST})$$

$$Catch \sim Tweedie(\text{Density}, \varphi^2, p)$$

$$\beta(\text{S}) \sim GMRF(0, \Sigma_1)$$

$$\beta(\text{ST}) \sim GMRF(0, \Sigma_2)$$

Survey catches are related to the underlying Cod density with a log-link function and a Tweedie distribution with variance and power parameters. The model included a fixed year effect representing annual averages, a fixed survey effect for the two data sources to account for gear differences (relative catch efficiency), a fixed depth effect using a cubic smoother, a spatial random effect ($\beta(\text{S})$), and a spatiotemporal effect ($\beta(\text{ST})$). The spatial effect can account for static spatial autocorrelation and is modelled by a Gaussian Random Markov Field (GMRF) where the covariance matrix, Σ_1, describes the pairwise covariance among locations depending on a de-correlation Matérn function (Lindgren et al. 2011). Likewise, the spatiotemporal effect models dynamic space-time auto-correlation using a GMRF between pairwise locations and years where the correlation across years is modeled with an autoregressive assumption (Yin et al. 2024). The model was implemented and optimized using the sdmTMB R-package (Anderson et al. 2022) built upon TMB (Kristensen et al. 2015).

Model diagnostics using the DHARMa residual diagnostics R-package (Hartig, 2022), where quantile residuals were calculated based on simulations from the fitted model, indicated sufficient fit (Figure 19), and residuals did not show significant spatial patterns (Figure 20) or bias related to the data sources (Figure 21). Results of estimated Cod biomass density distribution in 4X5Y during 2016–2024 indicate that the population did not exhibit significant large-scale changes during this period (Figure 22). Some localized fragmentation of previously continuous distribution appears in the deeper waters of Bay of Fundy, with Cod density decreasing from the area north of Jordan Basin, starting with 2019, despite being present there consistently throughout 2016–2018 (Figure 22). The population distribution featured a small contraction (Figure 24) and the estimated density-depth relationship indicated a preference at 50–200 m depth range throughout the whole time period (Figure 25). The overall population levels showed a gradual decrease until 2020, remaining stable at a low level thereafter, which is consistent with population trends from the 4X5Y assessment (Figure 23; DFO 2023a). The relative Cod density distribution was predicted to the bathymetry grid in preparation of bycatch analysis.

2.3. SUMMARY OF PREDICTORS

Candidate predictors for Cod bycatch are few, limited by both the AS program and the data format of CLFE. Recorded data from the AS program include exact fishing location and date; this spatial and temporal information can be used as is, or aggregated to various degrees to reduce noise, and more importantly, to bring in available environmental or fisheries data from

other sources (e.g., depth and Cod density as detailed above). Other recorded information such as trip, string, and trap identifiers, soak days and catches, or bycatch of other species may be able to inform Cod bycatch within a trap. For example, soak days have a defined relationship with Cod bycatch, while catches of Lobsters and other species may have an inverse relationship with Cod bycatch due to limited space in a trap. However, this information is missing from the reported CLFE data and hence, incorporating it in the analysis has very limited benefit to predicting total bycatch. A list of candidate predictors and a summary description are provided in Table 7. Overall, bycatch analyses are mainly limited by data availability (lack in multitude of potential predictors) and quality (resolution of data sampling).

3. METHOD

3.1. EXPLORATORY ANALYSIS

To inform model development and optimization, a preliminary analysis was conducted to explore the relationships between Cod bycatch and various potential predictors via data visualization. Cod bycatch within individual traps exhibit large variance (values of bycatch are highly dispersed in the original data). Models are likely unable to predict both accurately and precisely given observed distribution and considering difficulty to quantify the mechanism that results in Cod bycatch in a specific Lobster trap. Additionally, analysis is also limited by resolution of available data, i.e., most useful predictors in this study could not be differentiated to the level of an individual trap (e.g., Cod density, depth).

Data aggregation is an effective strategy to smooth out the noise and to explore variable correlations; the AS data were aggregated to the bathymetry grid so as to better investigate the following predictors: Cod density, depth, day of year (seasonal effect), and soak days (albeit absence from the CLFE data). Cod density (on the log scale) exhibits an expected strong, linear correlation with bycatch amounts (Figure 26) and supports the modeling results of this predictor. Relationship with depth indicates higher bycatch within the 50–150 m range, and possible, but uncertain (limited by sample sizes), depth preference within 150–350 m as well (Figure 27). Notably the mode (at around 100 m) and range in Figure 27 seem to deviate slightly from those in Figure 25 (at around 150 m), possibly suggesting that in deeper waters (>200 m) where Cod is likely scarce, bycatch does not reduce proportionally. Seasonality in Cod bycatch (Figure 28) is strong; during the winter/spring fishing seasons, bycatch is much higher. This partly corresponds to the different fishing seasons in different fishing areas as the fishing seasons for LFA 33 and 34 occur between late November to May 31 (day 1–150; 335–365), and possible seasonal distributional variations of 4X5Y Cod. Lastly, the relationship between Cod bycatch and soak days indicate a plateau at 7–9 days, but quickly tapers off after 10 days (Figure 29); this is either a saturation effect, or a possibility that Cod may have been consumed by Lobsters within the trap after a long period of time. This predictor was not included in the modeling stage as it is absent from the CLFE data. Though soak days could be imputed based on fishing dates and locations, model-estimated values likely contain duplicated information as these variables were already included in the bycatch analysis.

3.2. MEAN ESTIMATOR (ME)

The simplest model for bycatch analyses in data-limited scenarios is the mean estimator (Lohr, 2021), which we included as a benchmark for comparison across modeling frameworks. This simply represents the sample mean and variance of Cod bycatch from the AS data (observed data). Prediction for an unobserved trap is assumed to be the average bycatch weight from all observed traps within the same year. By extension, prediction of total bycatch is made based on total fishing effort. Residual diagnostics for this estimator exemplify the difficulty to predict to

individual trap-level bycatch, but when aggregated to the bathymetry grid, predictions were more reasonable (Figure 30).

Using this estimator to calculate the magnitude of total Cod bycatch makes the assumption that the observed data are independent and identically distributed (i.i.d.) random sub-samples from all commercial data and are thus representative of the latter. In reality, this assumption is likely not appropriate for many reasons: bycatch within close proximity is likely correlated; the trap-level observations are likely hierarchical within trips; operational constraints often influence the selection of observed traps and thus bias the sampling process. Sampling of AS traps is clearly biased for years before the at-sea sampling pilot program (DFO 2023b assessed potential bias within the pilot program for each LFA and suggested there was sampling bias within these years as well for some areas).

Most of the sampling bias is innate and difficult to amend via statistical modeling given limited knowledge of the underlying generating process. However, spatial and temporal correlations may be addressed to a certain extent to improve bycatch estimates either with statistical models, or simply by efficient data clustering (e.g., post-stratification and stratum-based prediction). During the model development stage in this study, various spatial clustering schemes were attempted to derive a stratified mean estimator instead of the simple mean estimator, and a 20-strata scheme based on K-means clustering ended up improving bycatch predictions. However, the improvement was minute compared to the following model-based approaches, and therefore not included in this document. This post-stratification scheme (Figure 31) was also tested as a potential predictor in some of the model-based approaches.

3.3. NON-PARAMETRIC MODEL

Random Forest proved to be an effective non-parametric model for bycatch analysis, especially in data-limited scenarios, e.g., bycatch analysis of multiple groundfish species in the United States west coast groundfish trawl fishery (Stock et al. 2019) and analysis of blue shark bycatch in the Canadian pelagic longline fishery (Yin et al., 2024). The Random Forest algorithm is a bootstrap aggregation of a large number of regression trees. Conceptually, a regression tree is similar to a stratified mean estimator since both methods group and stratify the data based on predictors. However, tree-based methods use a more general feature space of predictors of various types instead of physical space (i.e., spatial locations). They build a feature space from a set of candidate predictors where each predictor corresponds to a dimension (assuming non-collinearity), and partition the feature space into a set of rectangles where a simple regression model can be fit (Hastie et al. 2009). The tree is trained recursively to find the best partitioning regime to accurately predict the response variable, i.e. bycatch weight. Random Forest can be effective at building non-linear relationships between predictor and response variables (these relationships were assumed as piecewise flat). The “forest” relies on bootstrapping of randomized, independent decision trees, and hence, sometimes depends on large amounts of data to be effective.

The algorithm has three tuning parameters: the number of trees and two additional parameters related to randomization in bootstrapping and growth of the tree. A Boruta algorithm (Kursa and Rudnicki 2010) was used for pre-screening predictors where Random Forest classifiers were run repeatedly with a feature importance measure in order to evaluate and select a set of most important features (Breiman 2001, Hastie et al. 2009). We compared the importance scores calculated by Boruta and retained year, longitude, latitude, depth, day of year, and Cod density. Then the model was trained on the AS data, where tuning parameters were determined by cross-validation and a step-wise grid search (Figure 32). The R-package randomForest (Liaw et al. 2002) was used for model training and optimization. Diagnostics show small improvements

from the mean estimator (Figure 33) for positive catches, but predictions for zeros were not ideal.

3.4. GENERALIZED LINEAR MODELS

Generalized linear models (GLMs) is a framework encompassing a magnitude of statistical models that are widely used in ecology including simple linear regression models, generalized linear mixed models (GLMMs), generalized additive models (GAMs), etc. Contrary to Random Forest, these models are typically parametric or semi-parametric and assume specified and consistent relationships between predictor and response variables, i.e., implying pre-defined interactions. The relationships between the predictor and response variables are generalized from linear associations through either parametric functions (e.g., log-transformation) or non-parametric smoothing techniques (e.g., cubic smoothing). Model development in this framework followed the suggested workflow in Yin et al. (2024) to first select useful predictors and optimize their forms to be incorporated, and then progressing from simple and basic models to more complicated and nuanced formulas, evaluating model effectiveness, as well as, balance between complexity and predictive performance. After testing a variety of models (see Table 8 for a list of tested models and the model selection process), the best performing model is a GAM with the following specifications:

$$log(\mu) = \alpha(Year) + \mathrm{s}(\mathrm{Depth}) + \mathrm{s}(\mathrm{DOY}) + \log(\mathrm{Density}) + s(\mathrm{long}, \mathrm{lat})$$

$$Catch \sim Tweedie(\mu, \varphi^2, p)$$

Cod bycatch observations are assumed to follow a Tweedie distribution with mean, variance, and power parameters where the mean parameter is log-linked and predicted by year, depth, day of year, Cod density, and the spatial location. The GAM estimates dependency relationships using non-parametric smoothing techniques (cubic smoothing for depth and day of year; isotropic two-dimensional thin plate spline smoothing for spatial locations using longitude and latitude; see Wood 2011 for details), while year and density were included as categorical and continuous linear fixed effects, respectively. The GAM model was implemented using the mgcv R-package (Wood 2011) while other tested GLMs also used the glmmTMB package (Brooks et al. 2017) built upon TMB.

Model diagnostics (Figure 34) indicate that the Tweedie distribution assumption does not adequately represent the bycatch data. Given the amount of zeros (>98%) and sporadic nature of positive catches (extremely long-tailed), it is unlikely to find a better alternative of a parametric distribution. This model was selected as a candidate model for bycatch analysis despite this misrepresentation as its performance is acceptable in terms of predictive ability and in comparison with the benchmark ME. Even though predicting to individual traps is not ideal, predictions aggregated to the bathymetry grid indicate acceptable performance (Figure 35). In comparison with preliminary analysis, the optimized GAM reasonably captures the relationship with covariates (Figure 36).

3.5. SPATIAL MODEL

Similar to the spatial model for 4X5Y Cod density, a spatial model could be used to estimate Cod bycatch given assumptions of spatial and spatiotemporal autocorrelations of the assessed quantity. A spatial model can provide detailed predictions for any location based on spatial autocorrelation in the observations. Technically a spatial model is within the framework of GLMs where structured spatial and spatiotemporal random effects are explicitly modeled instead of a non-parametric approach in the GAM where a smoother over the two-dimensional space is used. Likewise to the GLM framework and the modeling of Cod density, we tested a range of different formulas pertaining to various underlying assumptions (see Table 9 for a list of tested

models and the model selection process), and the best performing model has the following specifications:

$$log(\mu) = \alpha(Year) + \log(\text{Density}) + s(\text{DOY}) + \beta(\text{S}) + \beta(\text{ST})$$

$$Catch \sim Tweedie(\mu, \varphi^2, p)$$

$$\beta(\text{S}) \ \sim GMRF(0, \Sigma_1)$$

$$\beta(\text{ST}) \sim GMRF(0, \Sigma_2)$$

Similar to the GAM, observations of Cod bycatch follows a Tweedie distribution where the mean parameter depends on year (fixed effect, discrete), Cod density (fixed effect, continuous and linear), day of year (using a cubic smoother), spatial and spatiotemporal effects (both structured random effects and defined similarly to the Cod density model in Section 2.2.2). Notably via model selection, depth was redundant in presence of structured spatial effects and Cod density was more favorable according to model evidence (Table 9). This is also supported by the preliminary analysis where predictability from Cod density appeared more defined than from depth.

Incorporation of Cod density undertook log-transformation that prevailed over alternatives during model development such as a linear form, a smoothing function or a linearly scaled transformation (values standardized to between 0 and 1 based on maximum and minimum). Eventually, the optimal model was equivalent to an assumption of linear catchability where catches were proportional to the underlying density, albeit variations dependant on space, time and seasonality. This interpretability further supported model selection and the presented model is optimal both mathematically and ecologically.

Model residual diagnostics and estimated relationships with predictors were similar to the GAM with potential mis-representation from the Tweedie distribution assumption (Figures 37-40). The two parametric models (GAM and STM) utilizing parametric distribution assumptions for the response variable were not able to represent the data characteristics with predominantly zero values and sporadic positive catches, and tended to over-predict zeros while underpredicting large values, which amounts to an averaging effect (Figures 35 and 38). ME was the best in representing zeros without any structural assumptions (Figure 30), but this performance could quickly discount facing sampling bias; RF were also better (Figure 33) due to lack of distribution assumptions, but also may not be robust. Performance of any estimators or models should be evaluated more rigorously (e.g., by predicting new data).

3.6. MODEL SELECTION

To evaluate and compare the four candidate models (ME, RF, GAM, and STM), a 5-fold cross-validation was used to compare the predictions from these candidate models and to prevent potential overfitting, which is likely to occur in models with complex structures. Cross-validation also assesses the reliability of a candidate model to predict “new” observations (additional observations that are not included in the model fitting phase), which is crucial to the objective of the bycatch analysis to predict unobserved traps given observed data. In a k-fold cross-validation, the partitioning of observations is subject to small random variations.

In 5-fold cross-validation, we randomly partitioned all observations into five groups. For each group (test subset), the model parameters (or tuning parameters in non-parametric models) were estimated using the other nine groups (training subset). The trained models were then used to predict on the test subset and residuals were calculated (test residuals). This training-testing process was iterated for all groups so that we had predictions and residuals for every observation. Cross-validated test residuals and residual variance were recorded, summarized

and compared across models. The same test and training subsets were used for all models to ensure results were comparable. Data partitioning used in cross-validation can exacerbate the problem of small sample sizes in data-limited scenarios and complicated models may be more sensitive to the cross-validation (Ding et al. 2018). Considering the small percentage of positive bycatch and to reduce the effect of randomization, we iterated the cross-validation 100 times using different random seeds in R. We used the root mean squared error (RMSE) and mean absolute error (MAE) as model evaluation metrics and averaged these metrics across all iterations. The use of the two metrics is equivalent to defining a statistical management risk combining both prediction accuracy and precision (Hastie et al. 2009). MAE penalizes large deviations more heavily, and thus, is more sensitive to outliers, and RMSE is more widely used as a balanced metric.

3.7. PREDICTING TOTAL BYCATCH

The best model from cross-validation was selected and used to predict annual total Cod bycatch. Predictions were also calculated using the mean estimator as a baseline for comparison (though it is important to emphasize that violations of assumptions can lead to wrong predictions using this method, especially for years 2016–2018 and 2024 where AS coverage were biased both spatially and temporally).

For LFAs 33–38, total trap hauls were aggregated and reported to the Lobster grids, while the model-based approaches (RF, GAM, and STM) were optimized using AS data and could predict to a higher spatial and temporal resolution than CLFE data. To address this misalignment, we took the following steps:

1. match the higher resolution bathymetry grid to the Lobster grid (based on grid centroids) and divide the Lobster grid using the bathymetry grid (Figure 41)
2. within each Lobster grid, evenly assign the total reported effort to each bathymetry grid (Figure 41); [Note: Since the bathymetry grid data available do not fully cover Lobster grid 322, the effort can be slightly over estimated. However, over or under-estimation of bycatch depends on effort distribution within this grid which in reality is unknown]
3. predict total Cod bycatch within each bathymetry grid based on Lobster fishing effort
4. for LFAs 33–38, find all grids within 4X5Y and sum up the total Cod bycatch

For LFA 41, locations and dates are available in CLFE data and thus predictions can be made for each set within the fishing area. There are 18 entries without any data and were removed (assuming errors and zero traps). For the remaining entries, proportions of missing data were calculated (Table 10); predictions were made for data without missing information first, and then scaled up with these proportions to account for missing data (i.e., imputed by predicted averages).

Predicted total Cod bycatch can be derived separately for each LFA and for all of 4X5Y by the above “divide-and-conquer” strategy. To evaluate prediction uncertainty, it is possible to calculate confidence intervals or standard deviations based on the Delta method in TMB (Kristensen et al. 2015) that accounts for inter-dependence among samples. However, the amount of observations (in the millions) would create a large covariance matrix that require exponential computation and proved to be unfeasible in this study. Instead, we calculated a standard deviation based on a simulation approach using sdmTMB, where the fitted model was used to simulate samples for the predictions. This uncertainty was calculated for inshore and offshore fisheries separately, as the two sets of data were in different formats, and subsequently combined assuming independence between the two (likely a mis-representation that over-projects uncertainty). Further, due to extremely long-tailed distribution of bycatch as a response

variable, and the fact that the parametric Tweedie distribution was likely a mis-representation, the simulation-based standard deviations are likely not accurate and should be interpreted with caution.

4. RESULTS

4.1. MODEL SELECTION RESULTS

Cross-validated residual RMSEs and MAEs were calculated for the four models in comparison, mean estimator (ME), random forest (RF), general additive model (GAM) and spatiotemporal model (STM). Using ME as a baseline, improvements to RMSE and MAE were calculated for each model and each year (Figures 42 and 43). The three model-based approaches all improved from the baseline and STM had the best performance; excluding years prior to the at-sea sampling pilot program (before 2018), improvements from STM were between 0.4%–1.8%. The unexplained variance pertaining to the individual trap-level variations probably were not representative of predictability for aggregated total Cod bycatch; in its stead, the cross-validated residuals were aggregated to the bathymetry grid (a coarser resolution) and a new set of RMSEs and MAEs were calculated (Figures 42 and 43; Tables 11 and 12). The STM performed the best, with improvements from the baseline ranging from 1–14% during the pilot program years. Using MAE as evaluation metric, STM still gave the best performance where reductions of aggregated MAE were around 20–35%. The spatial model prevailed at both approximating presence-absence (zeros) and predicting positive bycatch (Figure 38). The STM is recommended as the final model for Cod bycatch estimation.

4.2. PREDICTED TOTAL BYCATCH

Predicted annual total bycatch of Cod in the commercial Lobster fishery within 4X5Y is presented in Table 13 (numbers were aggregated to the entire study area; detailed numbers for each LFAs were not included for privacy protection). Predictions for 2024 used partial data and the predicted total does not include any fishing activities after October 2024; the analysis will be updated when all data become available and undergo quality control. Prior to the pilot program in 2016–2018, AS data coverage was scarce and total bycatch was mostly extrapolated based on commercial fishing efforts and estimated relationships between Cod bycatch and known predictors (spatial locations, seasonality, depth and Cod density) assuming consistency and stationarity of the underlying process. These numbers should be interpreted with caution as emphasized by the estimated uncertainty (standard deviation being multiples of estimates). Years 2020 and 2021 also had elevated uncertainty levels, possibly linked to sampling bias within these years due to COVID related restrictions (inconsistent sampling patterns during these years in Figures 7 and 8). In addition, predictions for 2017 are lower than ordinary due to the small number of AS data points in LFAs 33–38, all resulting in zero Cod bycatch. While for 2016, predictions relied on extrapolation solely as there were no AS data for these LFAs. Predictions for 2021 indicated high total Cod bycatch; this is mostly attributed to higher bycatch rates and in part, higher than ordinary fishing effort within LFA 36. In comparison to the baseline (estimated using the ME), STM mostly reduced predictions of bycatch and is believed to be a better estimate by accounting for spatial and temporal autocorrelations within the data. Overall, annual Cod bycatch amounts were in the range of 500–1000 metric tons. Given the consistency within AS observed Cod bycatch (Figure 9) and the consistency in percentage of observed traps (Table 10), these estimates were mostly in line with expectations.

Predictions of total Cod bycatch were summarized for each LFA (33–38 and 41) for management considerations (Table 14). The selected model (i.e., the spatial model) was able to predict relative distributions of Cod bycatch given a specific time frame, spatial area, and

presumed Lobster fishing effort; based on this, detailed spatial predictions of Cod bycatch for a few hypothetical scenarios were calculated and presented in Figures 44–46. The spatial information could be useful for future management planning to optimize resources while minimizing impact on Atlantic Cod stock in the region.

4.3. SOURCES OF UNCERTAINTY

In 2016–2018,prior to the pilot program, and in LFAs 36–38 across the study period, AS data coverage was scarce (Figure 6); therefore, total bycatch was mostly extrapolated based on commercial fishing efforts and estimated relationships between Cod bycatch and known predictors (spatial locations, seasonality, depth, and Cod density) assuming consistency and stationarity of the underlying process. These numbers should be interpreted with caution as is also emphasized by their estimated uncertainty (standard deviation being multiples of estimates). Years 2020 and 2021 also had elevated levels of uncertainty, possibly linked to sampling bias within these years due to COVID related restrictions (inconsistent sampling patterns during these years in Figures 7 and 8). Weighting risks and benefits, only predictions for 2018–2024 were recommended for management consideration.

The predictions are also subject to potential sampling bias related to operational limitations that were unable to be accounted for, and unexplained individual trap-level variance remain largely unaddressed due to data limitations. For example, the recommended model for bycatch analysis did not adequately represent the nature of bycatch observations (with overwhelming amount of zeros and sporadic positive catches). Uncertainty could also arise from simplifications in model assumptions, such as the absence of a seasonal effect in the Cod density model (predicted Cod density is based on surveys in the summer months due to data limitations therefore unable to account for distributional changes during fall/winter) and the separable seasonal effect in the bycatch model (assuming no interaction between seasonal and spatial effects). Further, calculation of confidence intervals (i.e., model-projected uncertainty around estimated total bycatch) were limited by computational powers and resorted to a simulation-based approach instead. This may be improved either by effective data aggregation prior to bycatch modeling or by streamlining the modeling process, although both would require further analysis and validation. In summary, the pilot program for at-sea sampling in the Lobster fishery can provide important information on Cod bycatch and would benefit from improved coverage, sampling consistency, and protocol compliance from an analytical perspective.

REFERENCES CITED

Anderson, S.C., Ward, E.J., English, P.A. and Barnett, L.A., 2022. sdmTMB: an R package for fast, flexible, and user-friendly generalized linear mixed effects models with spatial and spatiotemporal random fields. BioRxiv, pp.2022-03.

Andrushchenko, I., Clark, D., Irvine, F., MacEachern, E., Martin, R. and Wang, Y. 2022. 2018 2018 4X5Y Atlantic Cod Framework Data Inputs DFO Can. Sci. Advis. Sec. Res. Doc. 2022/006. v + 86 p.

Breiman, L., 2001. Random forests. Machine learning, 45, pp.5-32.

Brooks, M.E., Kristensen, K., Van Benthem, K.J., Magnusson, A., Berg, C.W., Nielsen, A., Skaug, H.J., Mächler, M. and Bolker, B.M., 2017. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling.

Cook, A. M., Denton, C., Zisserson, B. and Element, G. 2024. Maritimes Region American Lobster: Fisheries Spatial Distribution. Can. Tech. Rep. Fish. Aquat. Sci. 3602: iv + 121 p.

Cook, A.M., Denton, C., Hubley B., and Cassista-Da Ros, M. In prep. Inshore Lobster trawl survey: Development of trawl conversion factors and the importance of trawl mensuration data. Tech. Rep. Fish. Aquat. Sci. XXX: v + # p.

Criquet G., Brêthes J-C, and Allain RJ. 2015. Marine Stewardship Council Assessment: Public Certification Report for the Nova Scotia and New Brunswick Lobster Eco-Certification Society. Bay of Fundy, Scotian Shelf and Southern Gulf of St. Lawrence Lobster (Homarus americanus) Trap Fisheries. MSC 013.

Denton, Cheryl M. 2020. Maritimes Region Inshore Lobster Trawl Survey Technical Description. Can. Tech. Rep. Fish. Aquat. Sci. 3376: v + 52 p.

DFO. 2013. Policy on Managing Bycatch. Sustainable Fisheries Framework.

DFO. 2021. Rebuilding plan for Atlantic cod – NAFO Division 4X5Y.

DFO. 2023a. Stock Status Update of Atlantic Cod (*Gadus morhua*) In NAFO Divisions 4X5Y for 2022v. DFO Can. Sci. Advis. Sec. Sci. Resp. 2023/017.

DFO. 2023b. Bycatch Analyses from the Inshore Lobster Fisheries in Lobster Fishing Areas 27, 31a, 31b, 33, 34, and 35. DFO Can. Sci. Advis. Sec. Sci. Advis. Rep. 2023/032.

DFO. 2024. Maritimes Research Vessel Survey Trends on the Scotian Shelf and Bay of Fundy for 2023 DFO Can. Sci. Advis. Sec. Sci. Resp. 2024/010.

Ding, J., Tarokh, V. and Yang, Y., 2018. Model selection techniques: An overview. IEEE Signal Processing Magazine, 35(6), pp.16-34.

Hartig, F., 2016. DHARMa: residual diagnostics for hierarchical (multi-level/mixed) regression models. CRAN: Contributed Packages.

Hastie, T., Tibshirani, R., Friedman, J.H. and Friedman, J.H., 2009. The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758). New York: springer.

Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H. and Bell, B.M., 2016. TMB: automatic differentiation and Laplace approximation. Journal of statistical software, 70, pp.1-21.

Kursa, M.B. and Rudnicki, W.R., 2010. Feature selection with the Boruta package. Journal of statistical software, 36, pp.1-13.

Liaw, A. and Wiener, M., 2002. Classification and regression by randomForest. R news, 2(3), pp.18-22.

Lindgren, F. and Rue, H., 2015. Bayesian spatial modelling with R-INLA. Journal of statistical software, 63, pp.1-25.

Lindgren, F., Rue, H. and Lindström, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. Journal of the Royal Statistical Society Series B: Statistical Methodology, 73(4), pp.423-498.

Pezzack, D.S., Denton, C.M., and Tremblay, M. J. 2014. Overview of By-catch and Discards in the Maritimes Region Lobster Fishing Areas (LFAs) 27-33 based on Species at Risk Act (SARA) At-sea Sampling 2009-2010 DFO Can. Sci. Advis. Sec. Res. Doc. 2014/040. v + 27 p.

Stock, B.C., Ward, E.J., Thorson, J.T., Jannot, J.E. and Semmens, B.X., 2019. The utility of spatial model-based estimators of unobserved bycatch. ICES Journal of Marine Science, 76(1), pp.255-267.

Treguier, A.M., Barnier, B., Albert, A., Deshayes, J., leSommer, J., Lique, C., Molines, J.M., Penduff, T. and Talandier, C., 2017, April. ORCA12, a global ocean-ice model at 1/12°: successes, shortcomings and their impact on ocean forecasting. In EGU General Assembly Conference Abstracts (p. 17950).

Wang, Z., Lu, Y., Greenan, B., Brickman, D., and DeTracey, B., 2018. BNAM: An eddy resolving North Atlantic Ocean model to support ocean monitoring. Can. Tech. Rep. Hydrogr. Ocean. Sci. 327: vii + 18p.

Wood, S.N., 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society Series B: Statistical Methodology, 73(1), pp.3-36.

Yin, Y., Bowlby, H.D. and Benoît, H.P., 2024. A roadmap for generating annual bycatch estimates from sparse at-sea observer data. ICES Journal of Marine Science, 81(9), pp.1850-1867.

Yin, Y., Benoît, H., and Martin, R. 2025. Calibration of Bottom Trawl Survey Vessels: Results of Comparative Fishing Between the CCGS Teleost and CCGS John Cabot/ CCGS Captain Jacques Cartier on the Scotian Shelf and Bay of Fundy in 2022 and 2023 DFO Can. Sci. Advis. Sec. Res. Doc. 2025/025. iv + 236 p.

FIGURES

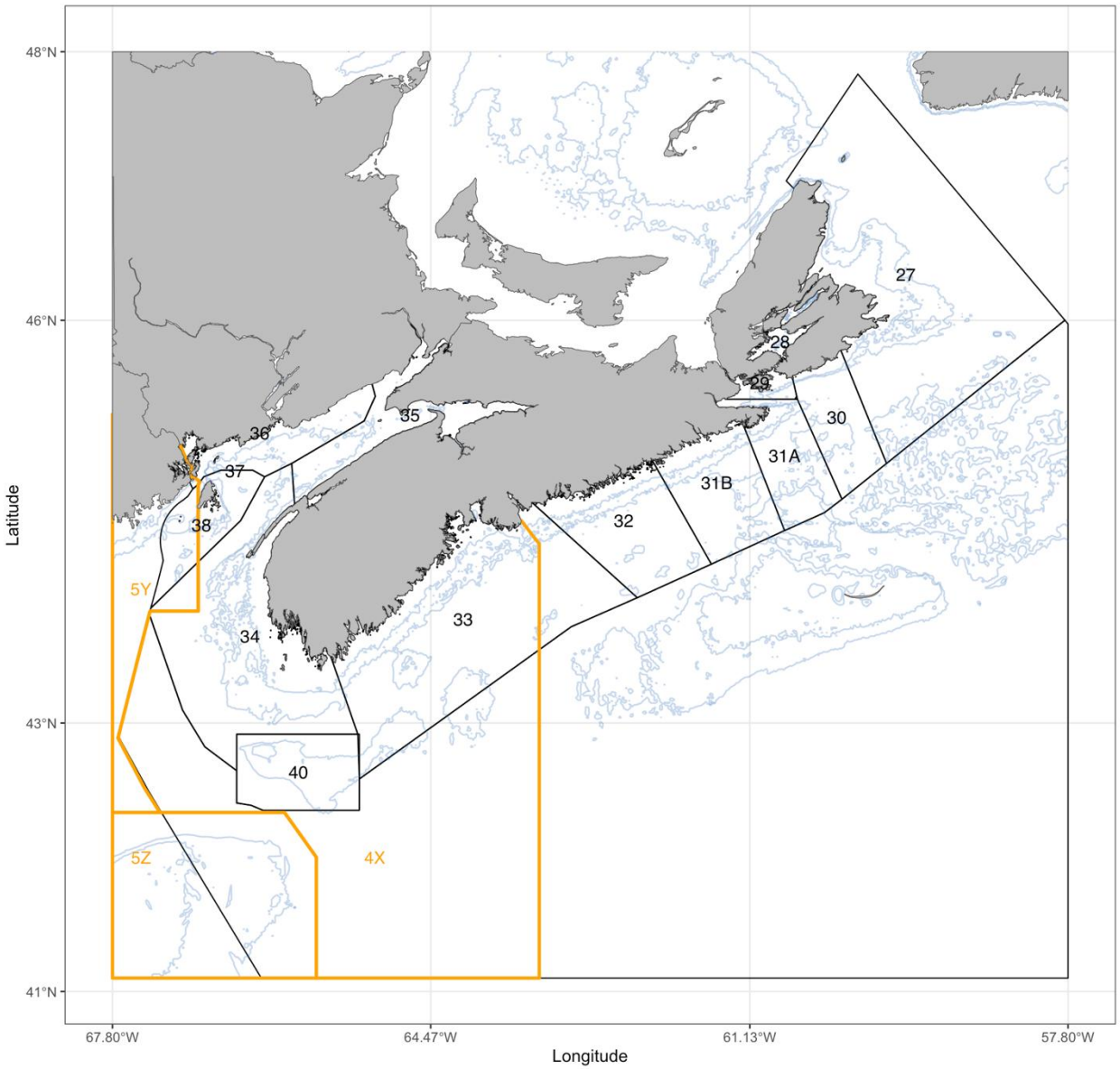


Figure 1. Map of DFO Maritimes Region lobster fishing areas (black lines) and Northwest Atlantic Fisheries Organization Divisions 4X5Y (yellow lines). The blue lines represent depth contours.

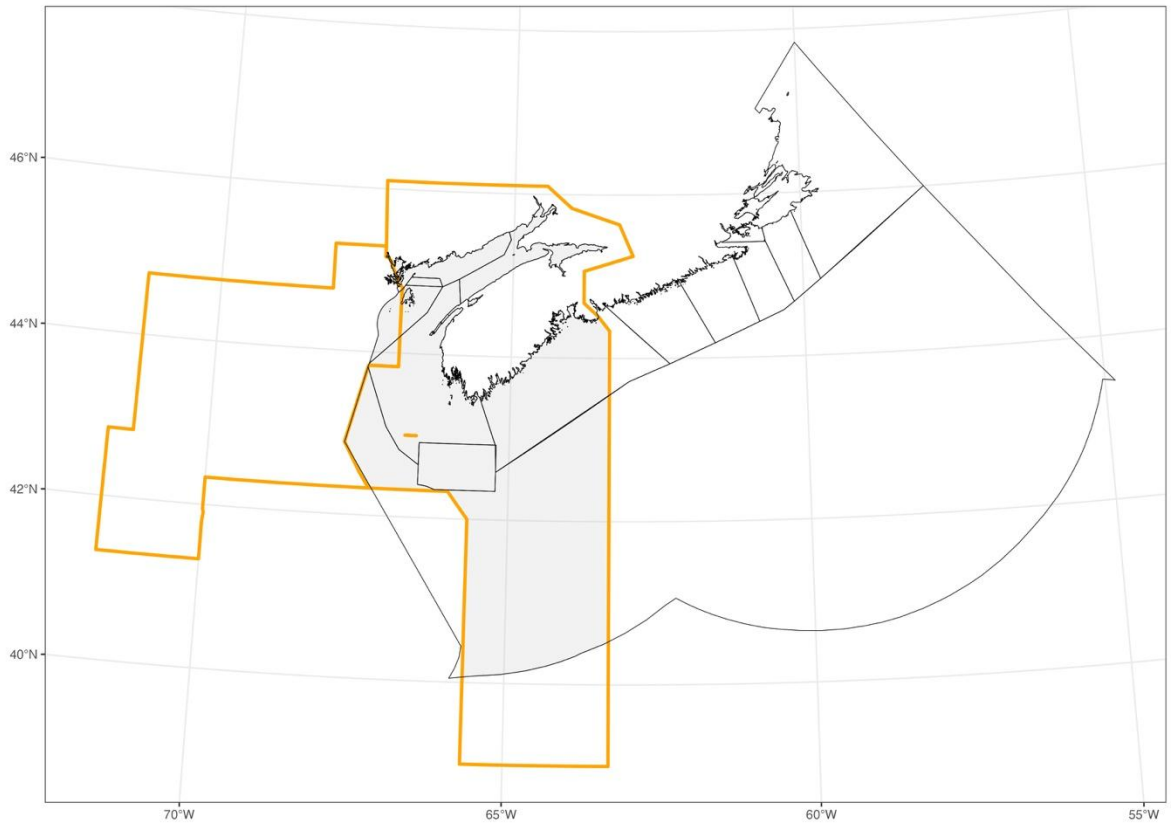


Figure 2. The study area (shaded in gray) is the overlap between Northwest Atlantic Fisheries Organization Divisions 4X5Y (yellow lines) and Maritimes Region Lobster Fishing Areas (black lines) where cod bycatch in the commercial lobster fishery is of interest.

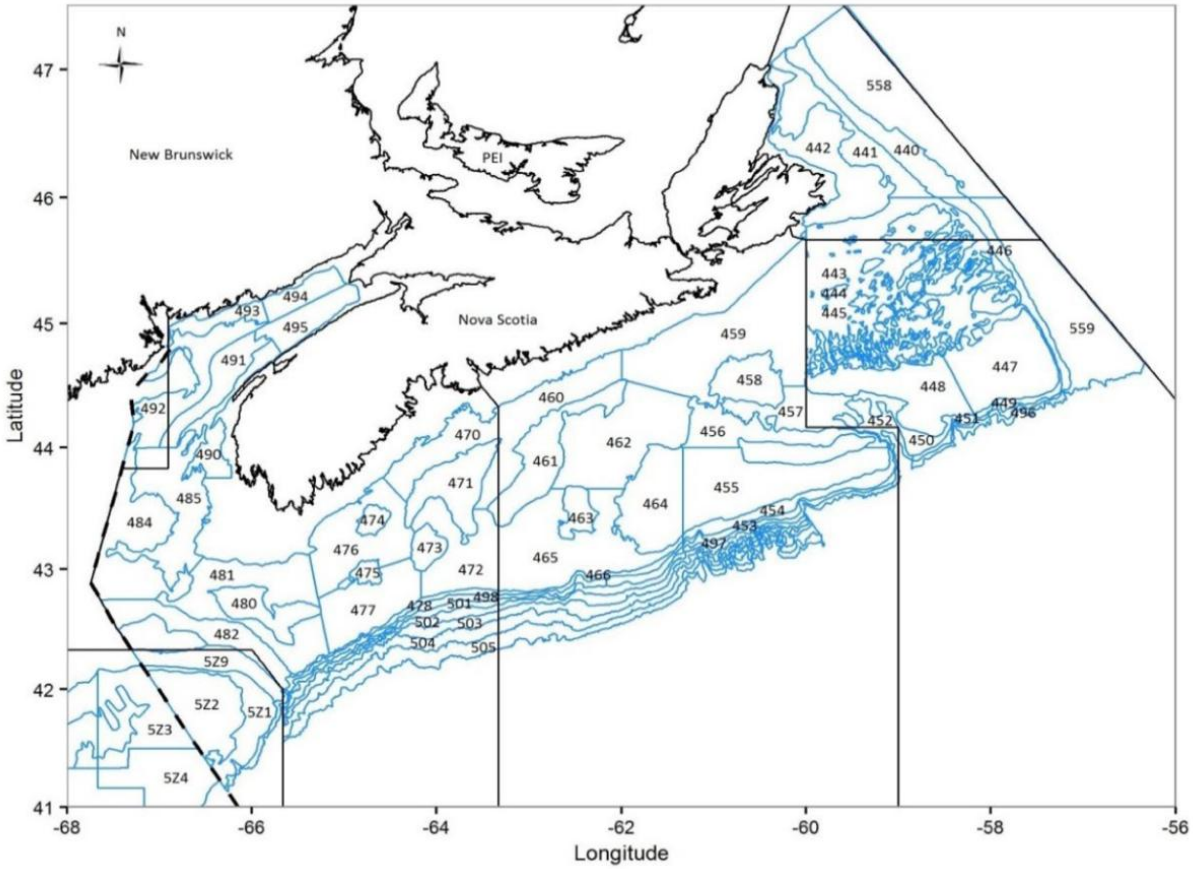


Figure 3. DFO Summer Ecosystem Research Vessel Survey strata (DFO 2024).

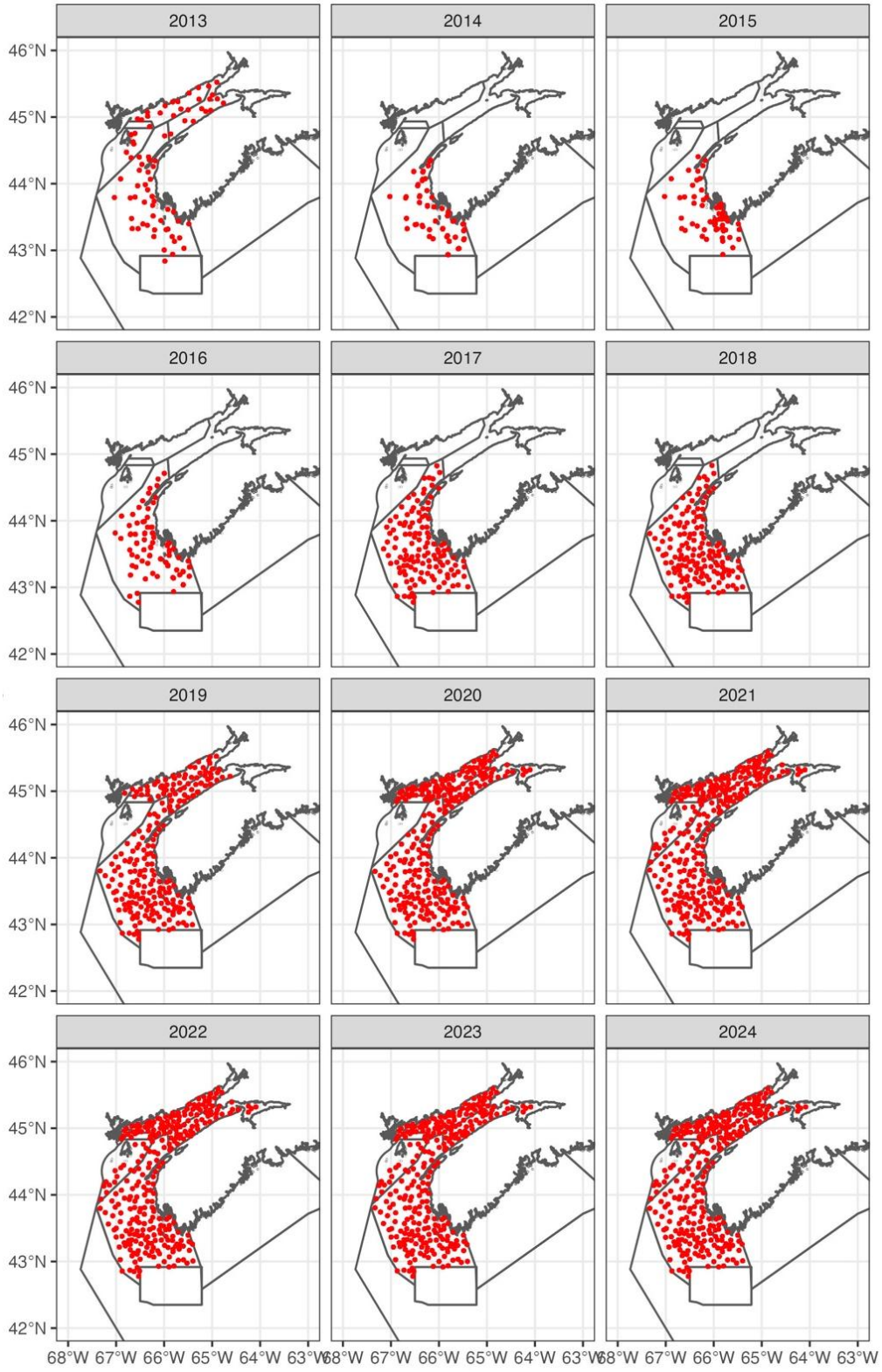


Figure 4. Map of fixed station locations of the Inshore Lobster Trawl Survey from 2013 to 2023 in June and July. Red points represent locations of survey tows.

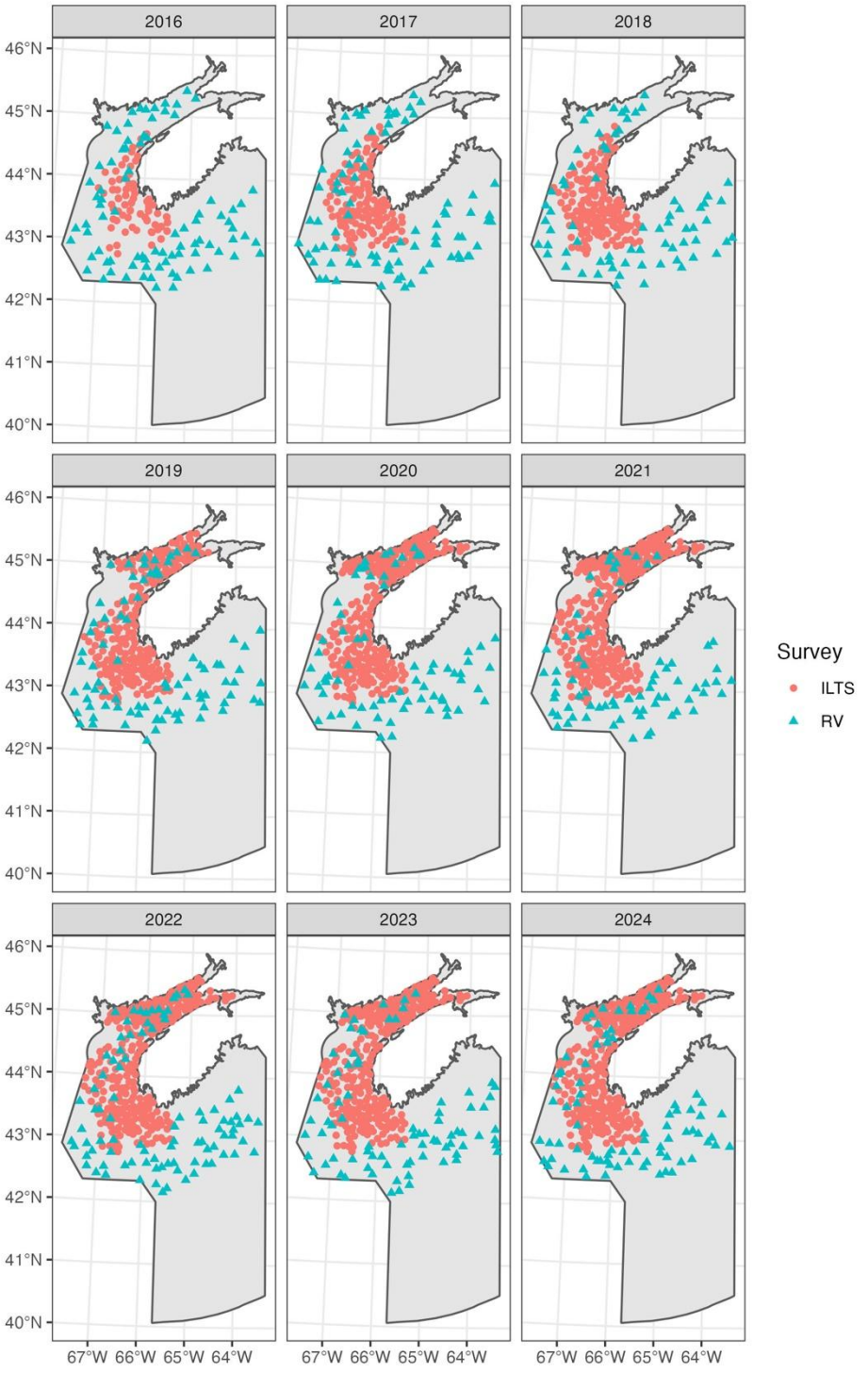


Figure 5. Locations of the DFO Summer Ecosystem Research Vessel Survey (RV) and the Inshore Lobster Trawl Survey (ILTS) within the study area and during 2016-2024. The RV survey has random stations and ILTS has fixed stations.

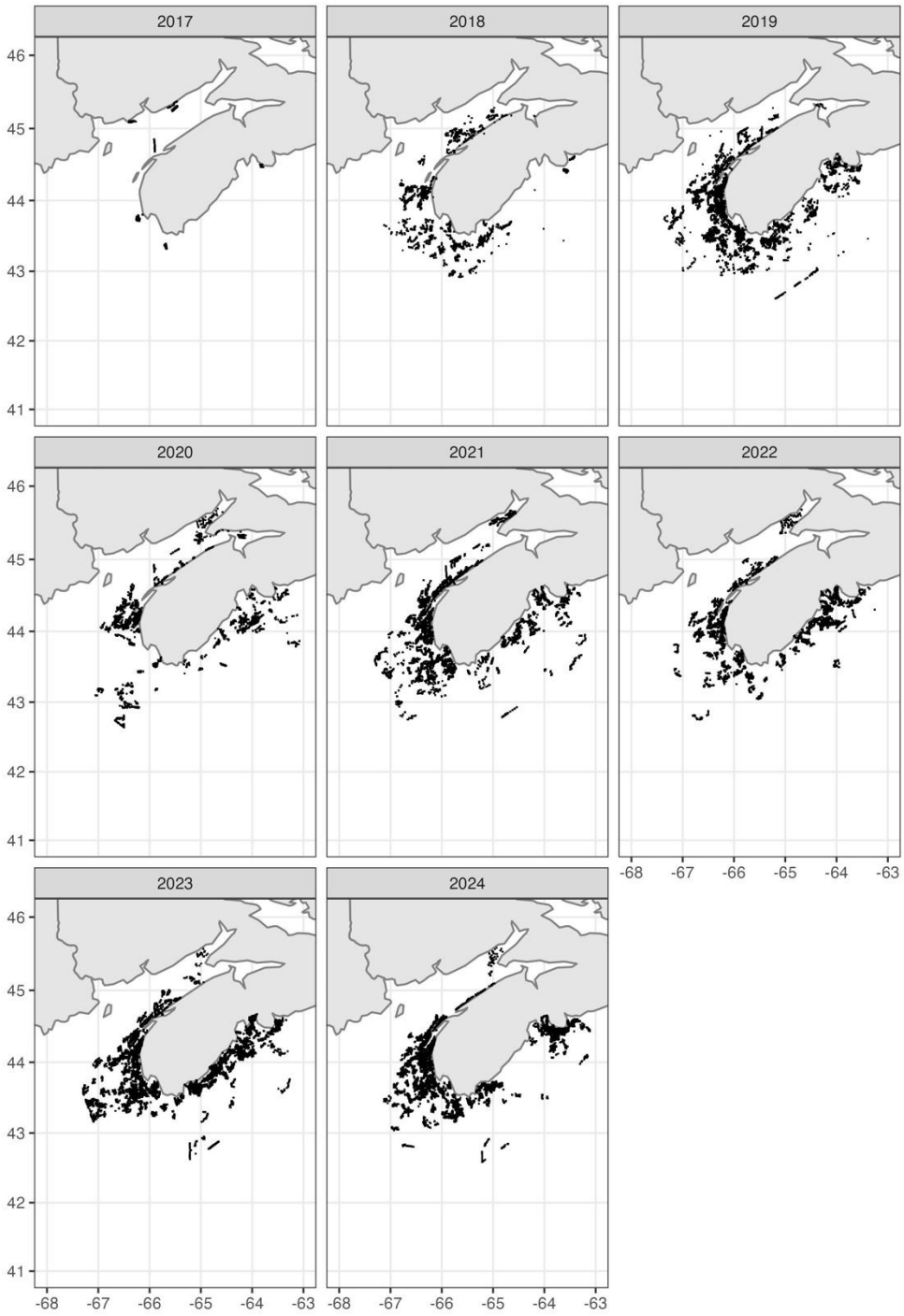


Figure 6. Spatial coverage of the at-sea Sampling data (locations where cod bycatch was monitored). Locations in lobster fishing area 41 were removed for privacy protection. All observed locations in 2016 were in LFA 41.

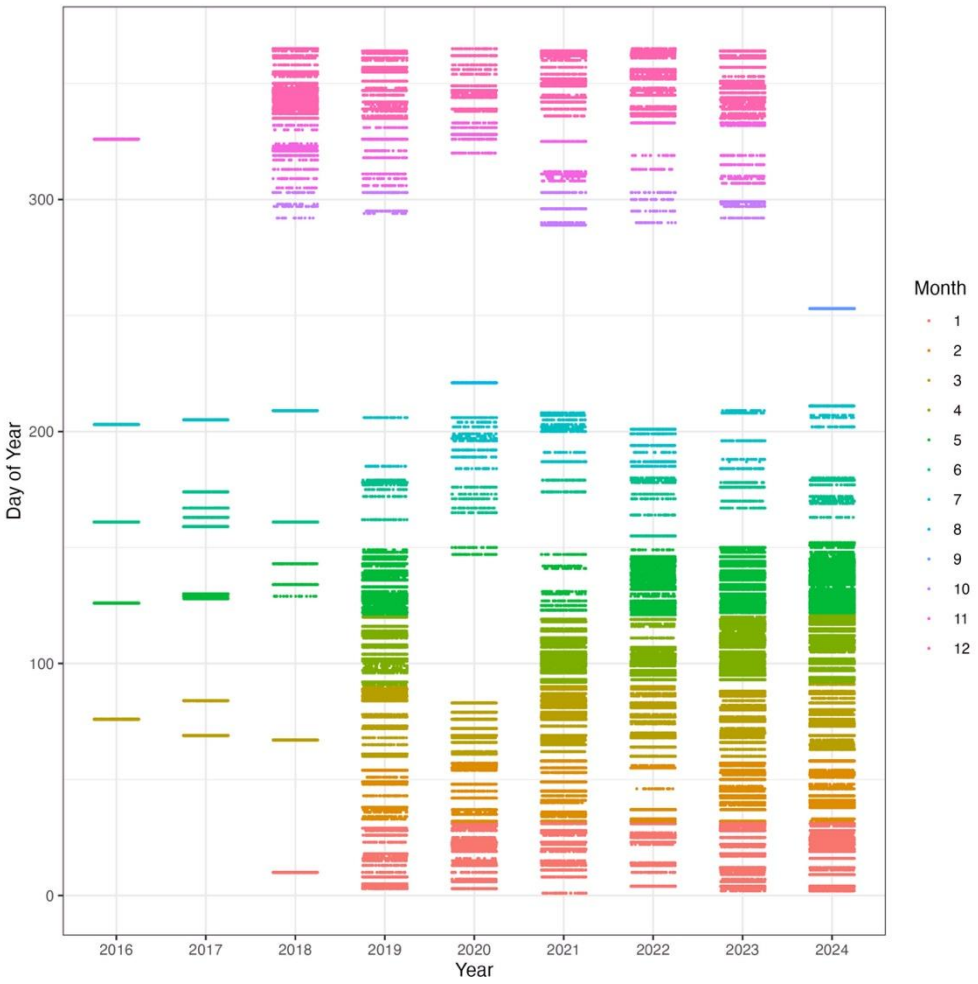


Figure 7. Temporal coverage of the at-sea sampling data (day of year).

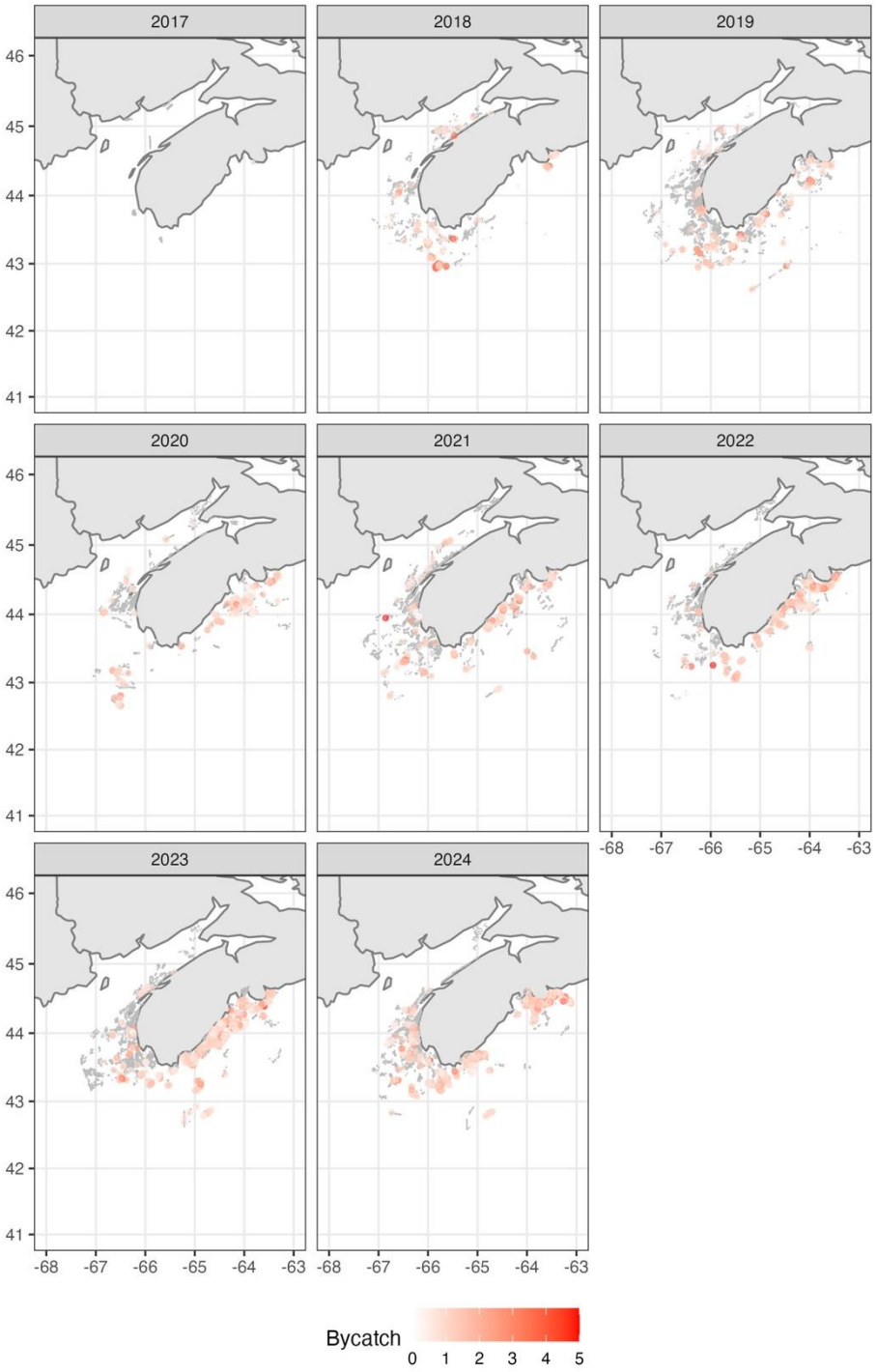


Figure 8. Cod bycatch in the at-sea sampling data where color intensity indicates catch weights (kg) and gray dots indicate zero bycatch. Locations in lobster fishing area 41 were removed for privacy protection. All observed locations in 2016 were in LFA 41.

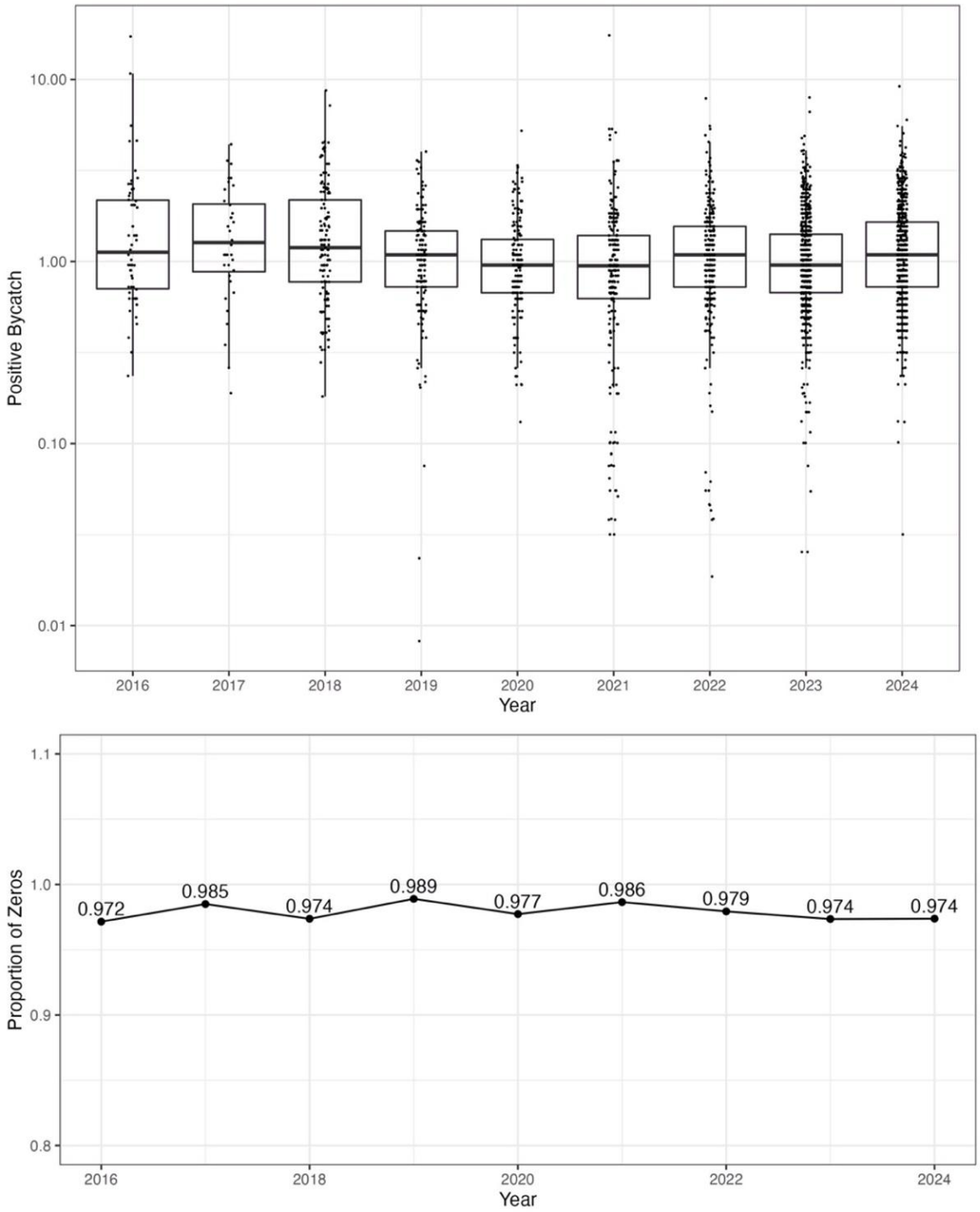


Figure 9. Summary of cod bycatch in the at-sea sampling data: annual trend of positive catches and proportion of zeros.

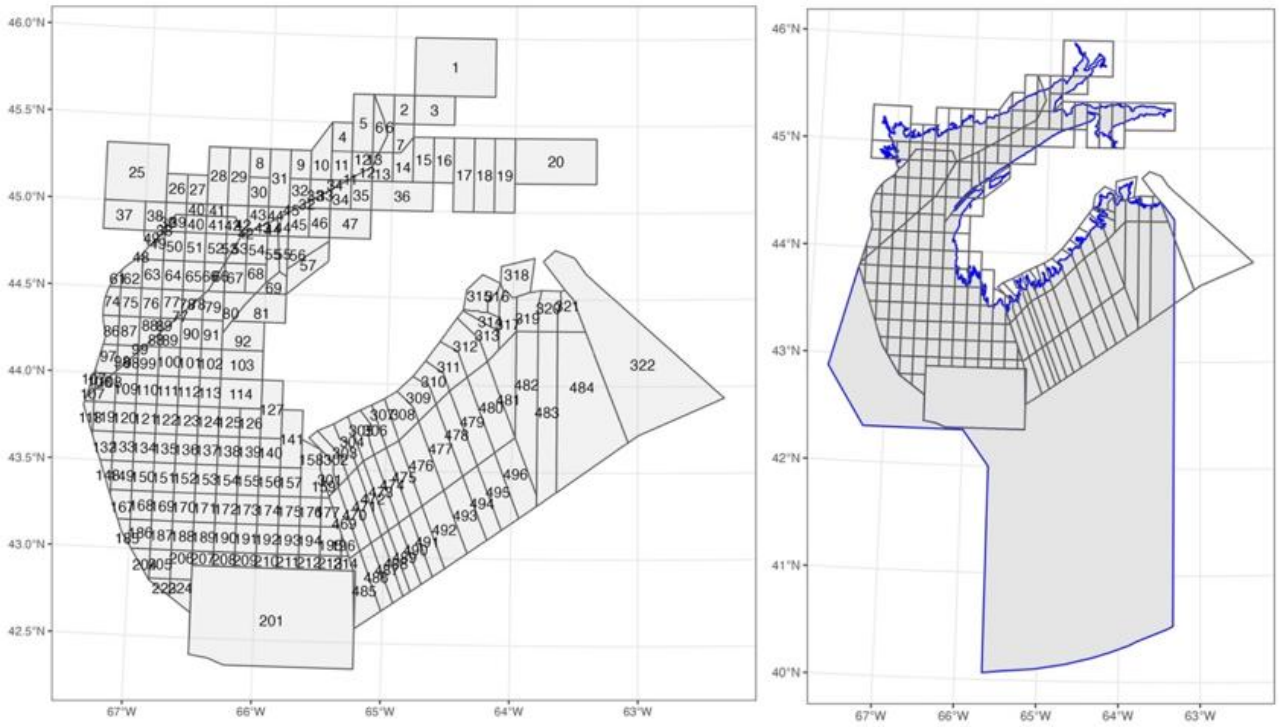


Figure 10. Reporting grid (left panel) for commercial inshore lobster fishing effort data, i.e., lobster grid, and the overlap between the lobster grid and the study area (right panel, shaded in gray).

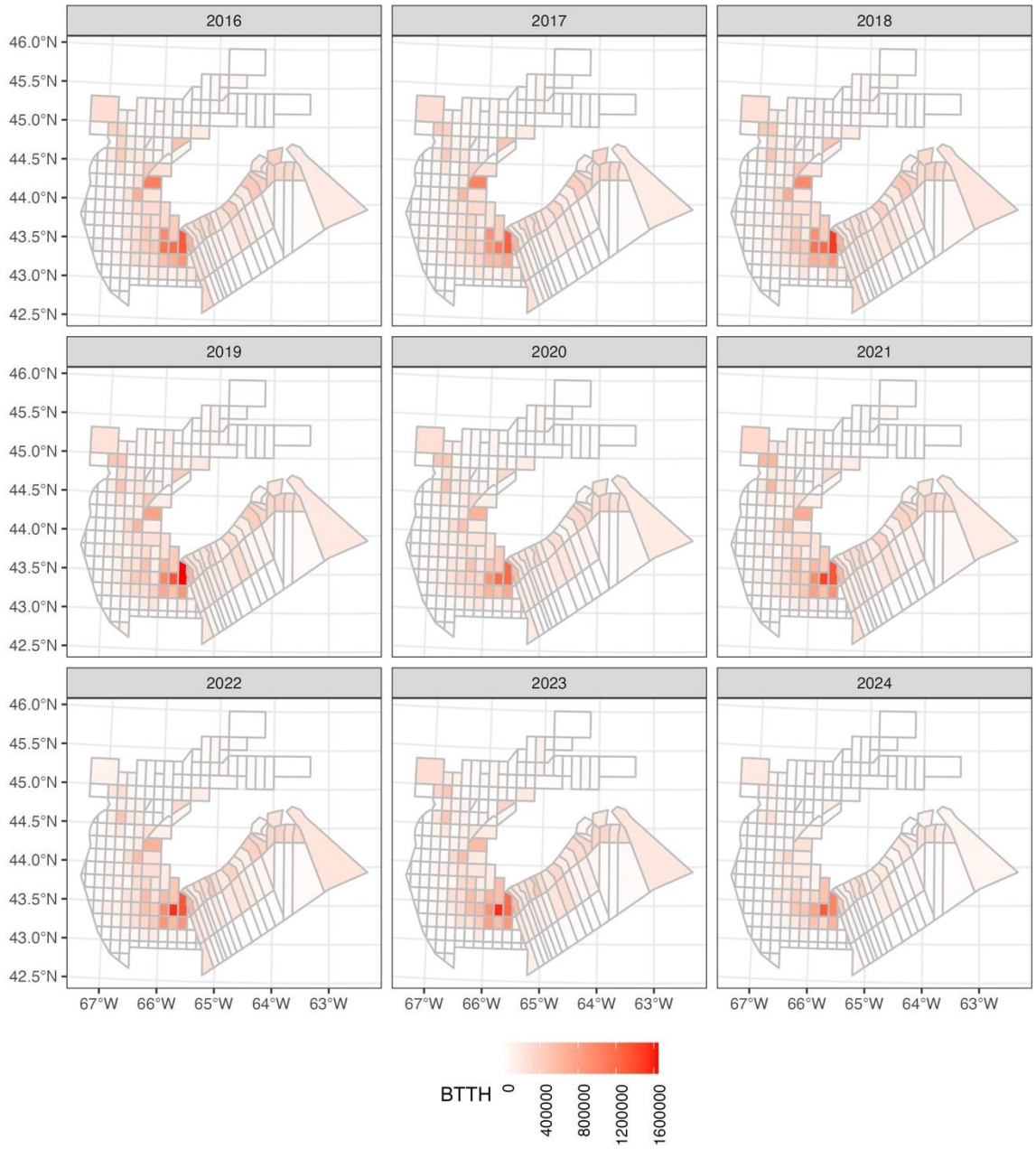


Figure 11. Spatial distribution of the commercial lobster fishing effort for inshore lobster fishing areas (LFAs) related to the study area where BTTH (bumped-up total trap hauls) is the estimated total number of traps within each lobster grid. Data for LFA 41 were not shown in the figure for privacy protection.

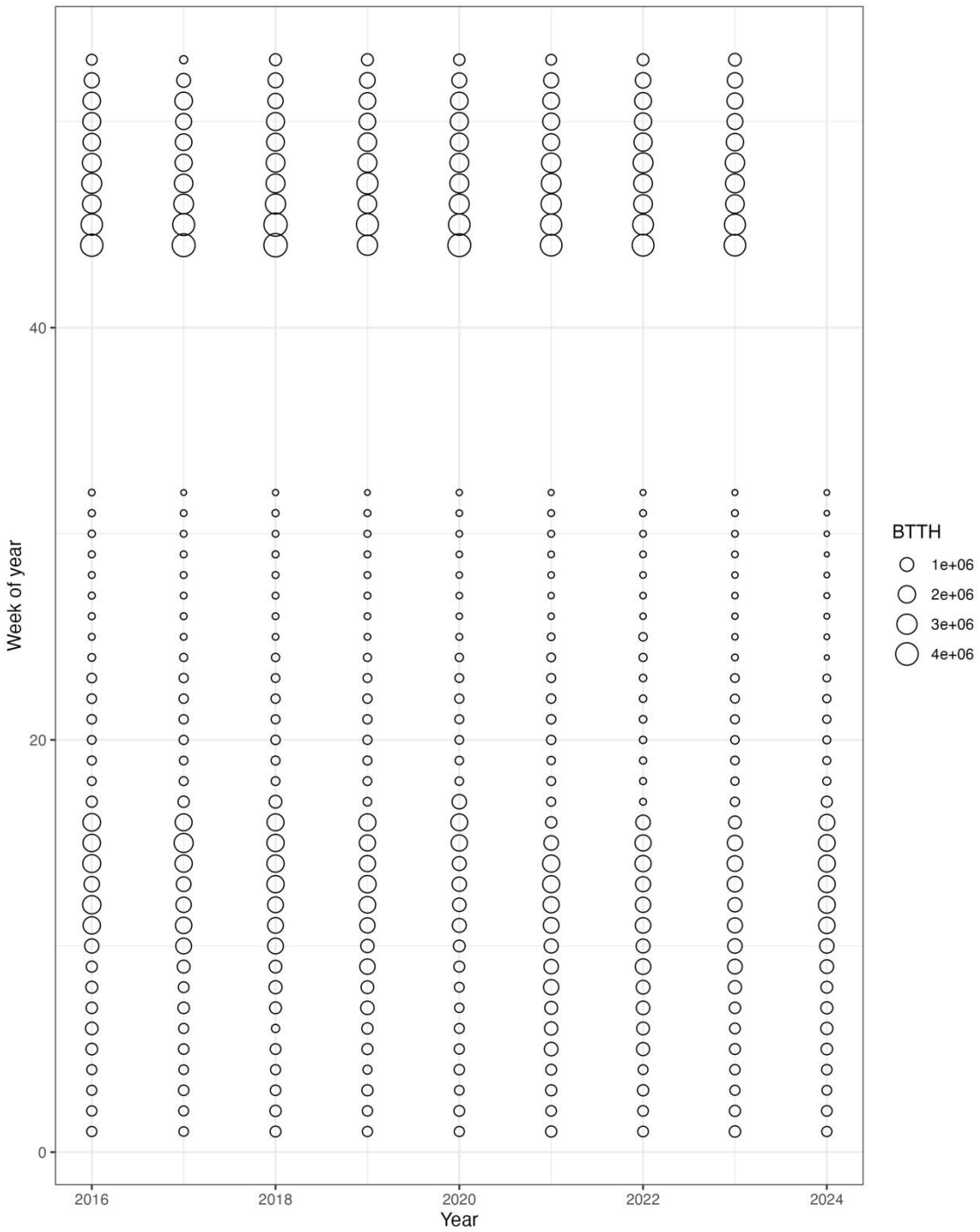


Figure 12. Temporal distribution of the commercial lobster fishing effort for inshore LFAs (lobster fishing areas) related to the study area where BTTH (bumped-up total trap hauls) is the estimated total number of traps. Data for LFA 41 were removed from the figure for privacy protection.

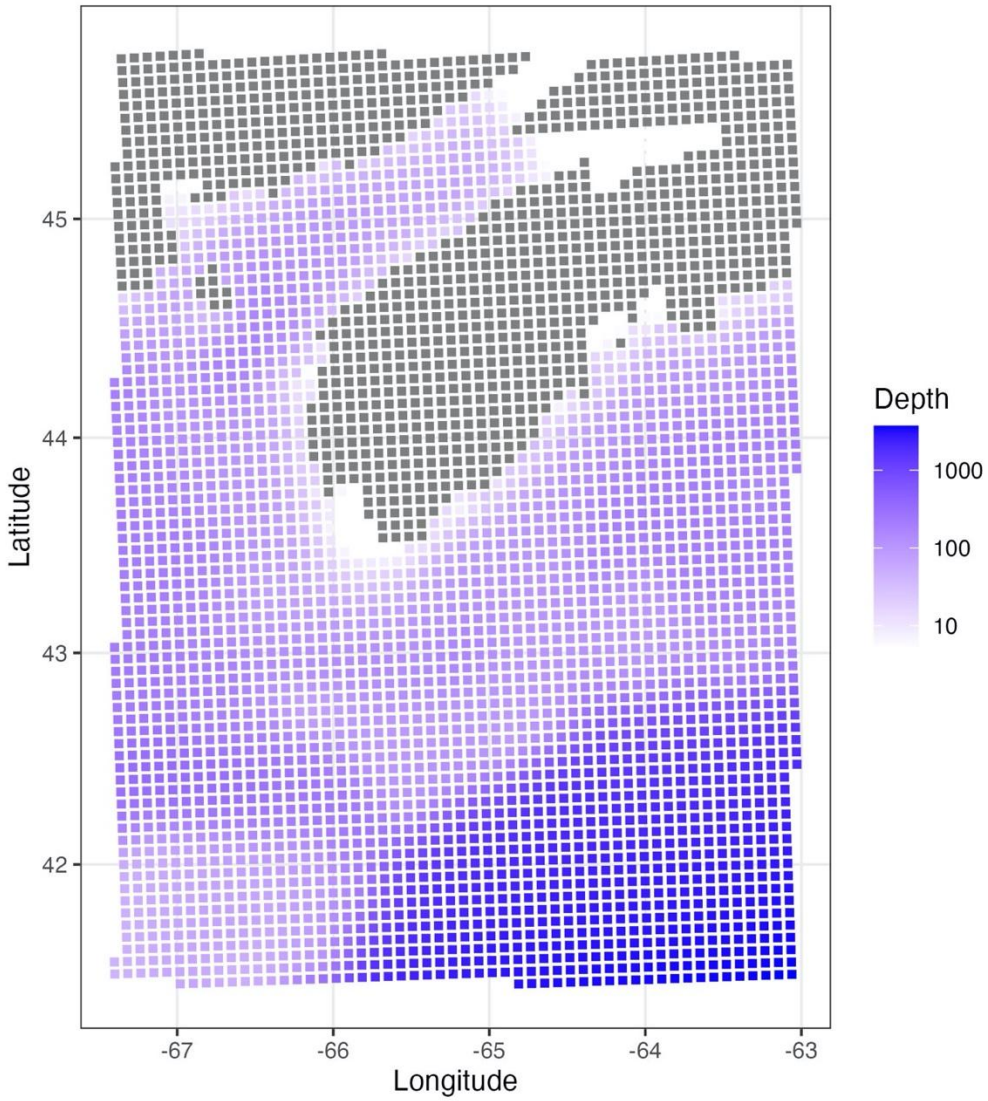


Figure 13. Bathymetry data and grid used in this study (grey cells are land).

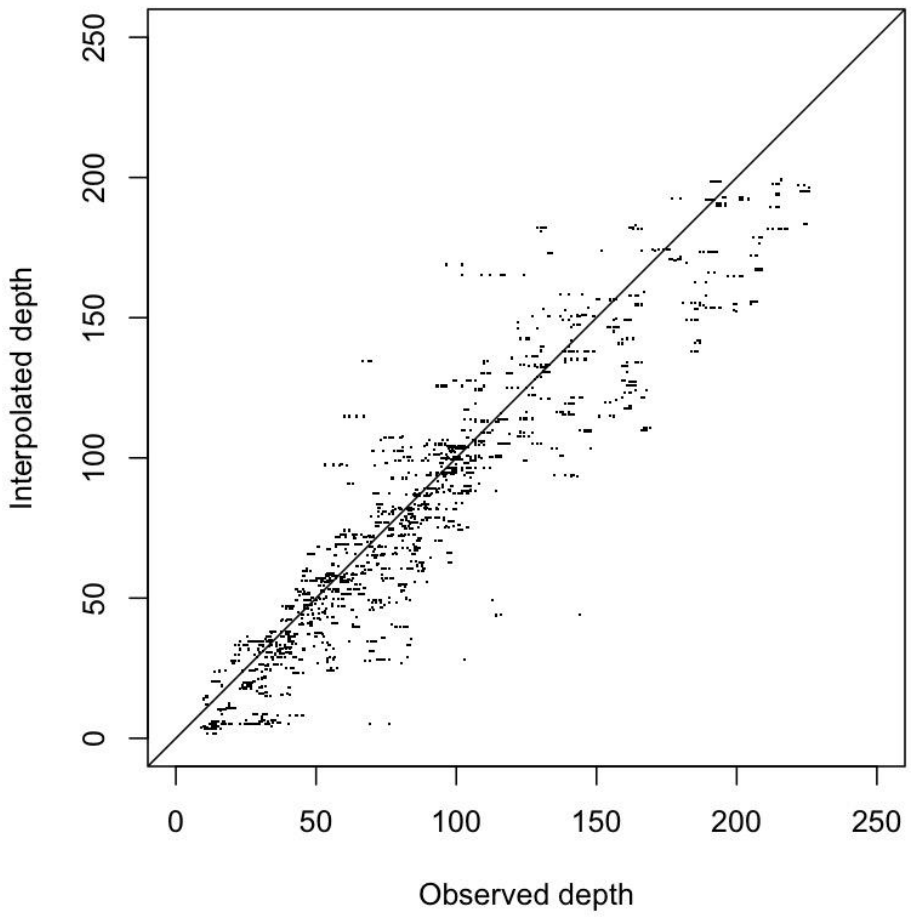


Figure 14. Difference between observed depth from the Inshore Lobster Trawl Survey trips and the high quality depth data used in this study (interpolated to the observation locations using a 5km radius and averaging).

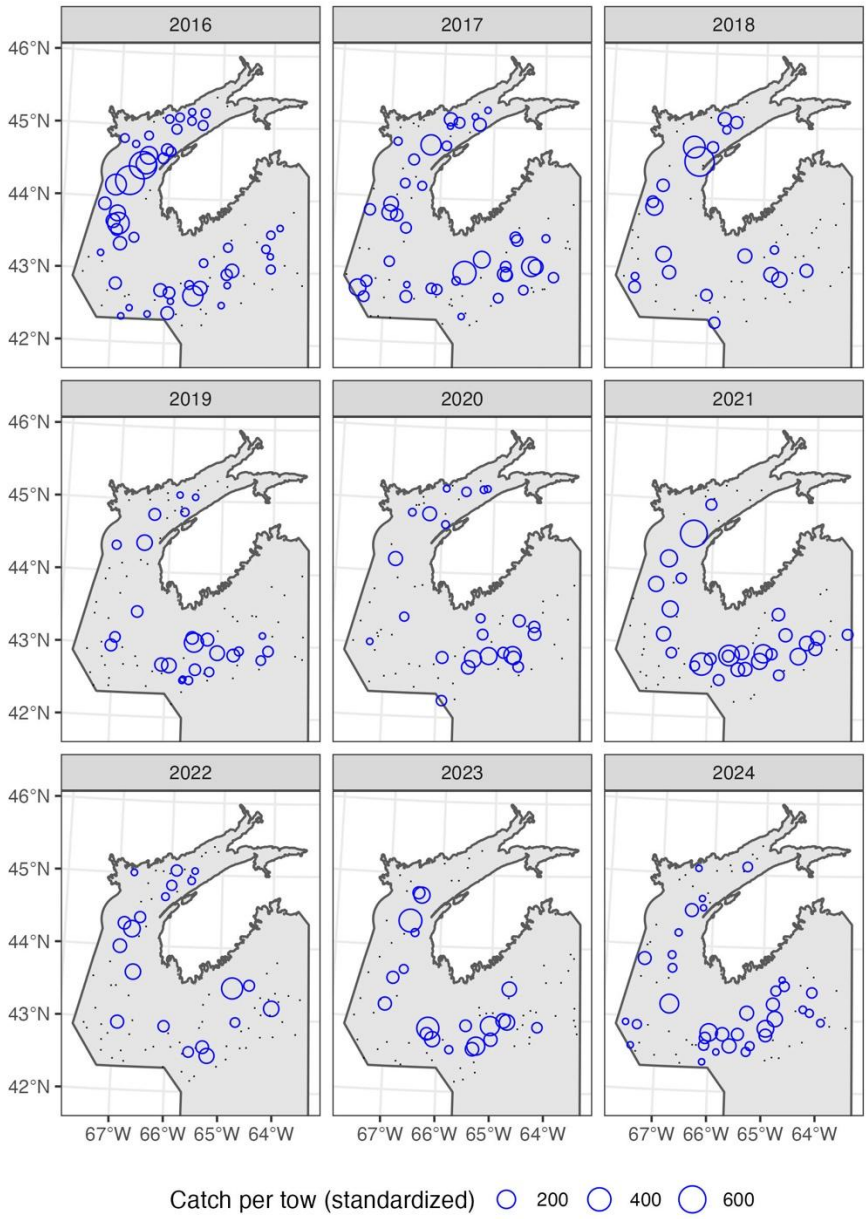


Figure 15. DFO Summer Ecosystem Survey (RV Survey) data: catch weights per tow for cod (circle size indicate catch values and dots indicate zero catches). Catches were standardized for tow distance and assuming a trawl opening of 14ft; historical gear changes were calibrated incorporating conversion factors from comparative fishing analyses.

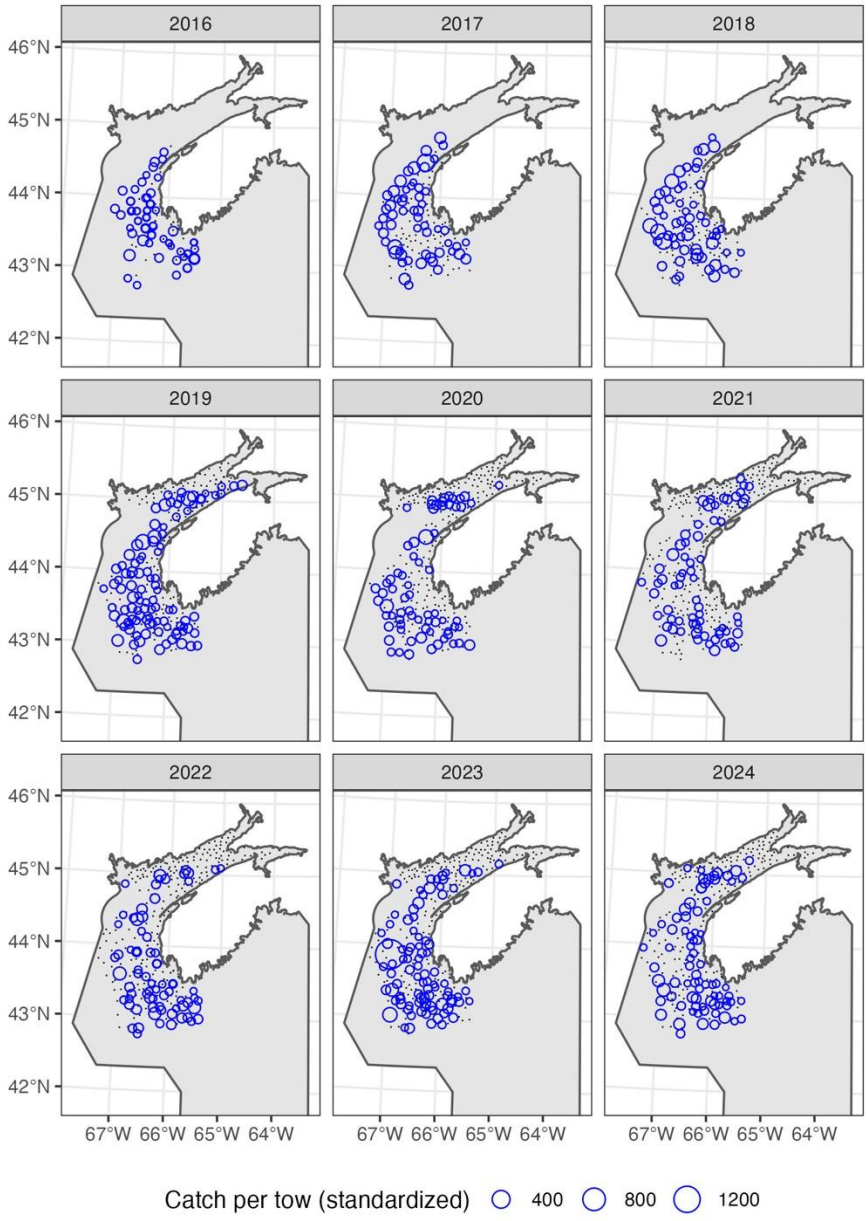


Figure 16. Inshore Lobster Trawl Survey data: catch weights per tow for cod (circle size indicate catch values and dots indicate zero catches). Catches were standardized for area towed; historical gear changes were calibrated incorporating conversion factors from comparative fishing analyses.

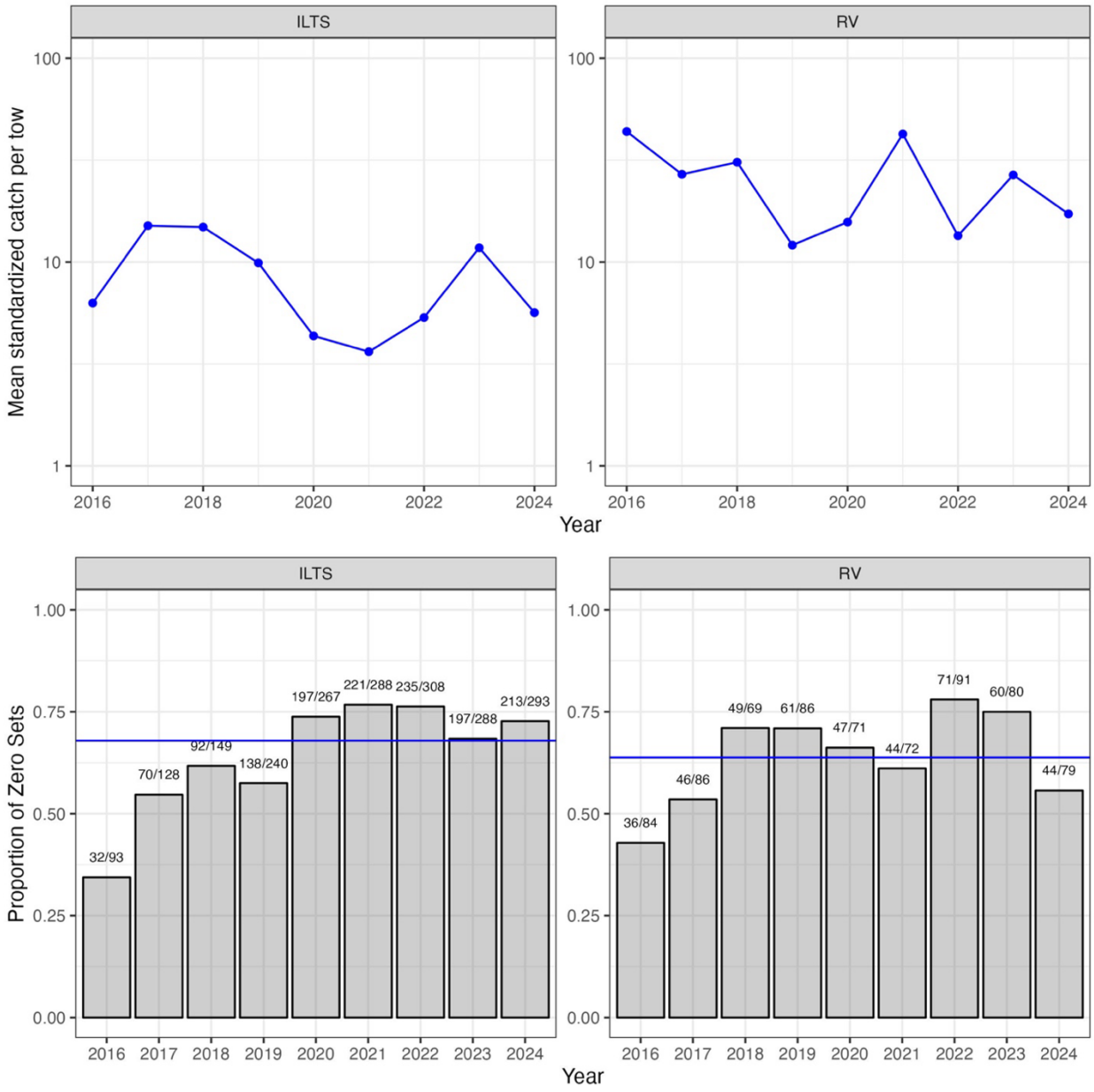


Figure 17. Summary of cod catch from the DFO Summer Ecosystem Survey and Inshore Lobster Trawl Survey.

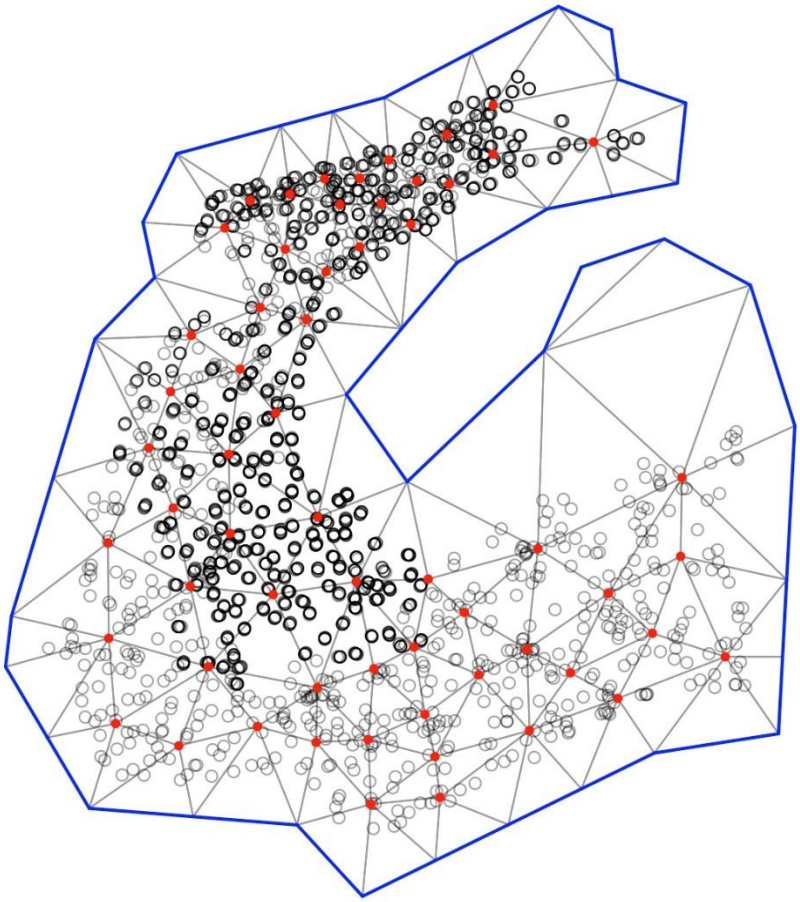


Figure 18. Area boundary (blue lines) and the INLA mesh for cod density modeling. Gray points are station locations from DFO Summer Ecosystem Survey and black points are locations from Inshore Lobster Trawl Survey; red points are the "knots" upon which a spatial covariance matrix was built for the spatial model.

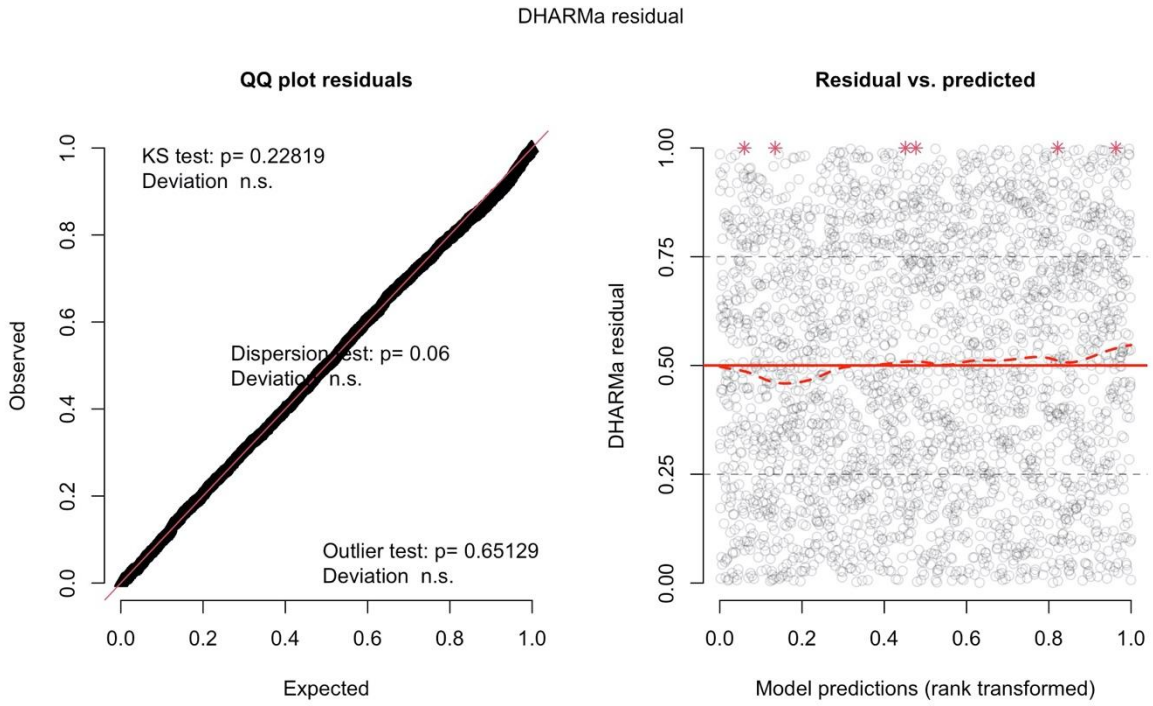


Figure 19. Model diagnostics using the DHARMa residual diagnostics R package indicate sufficient fit.

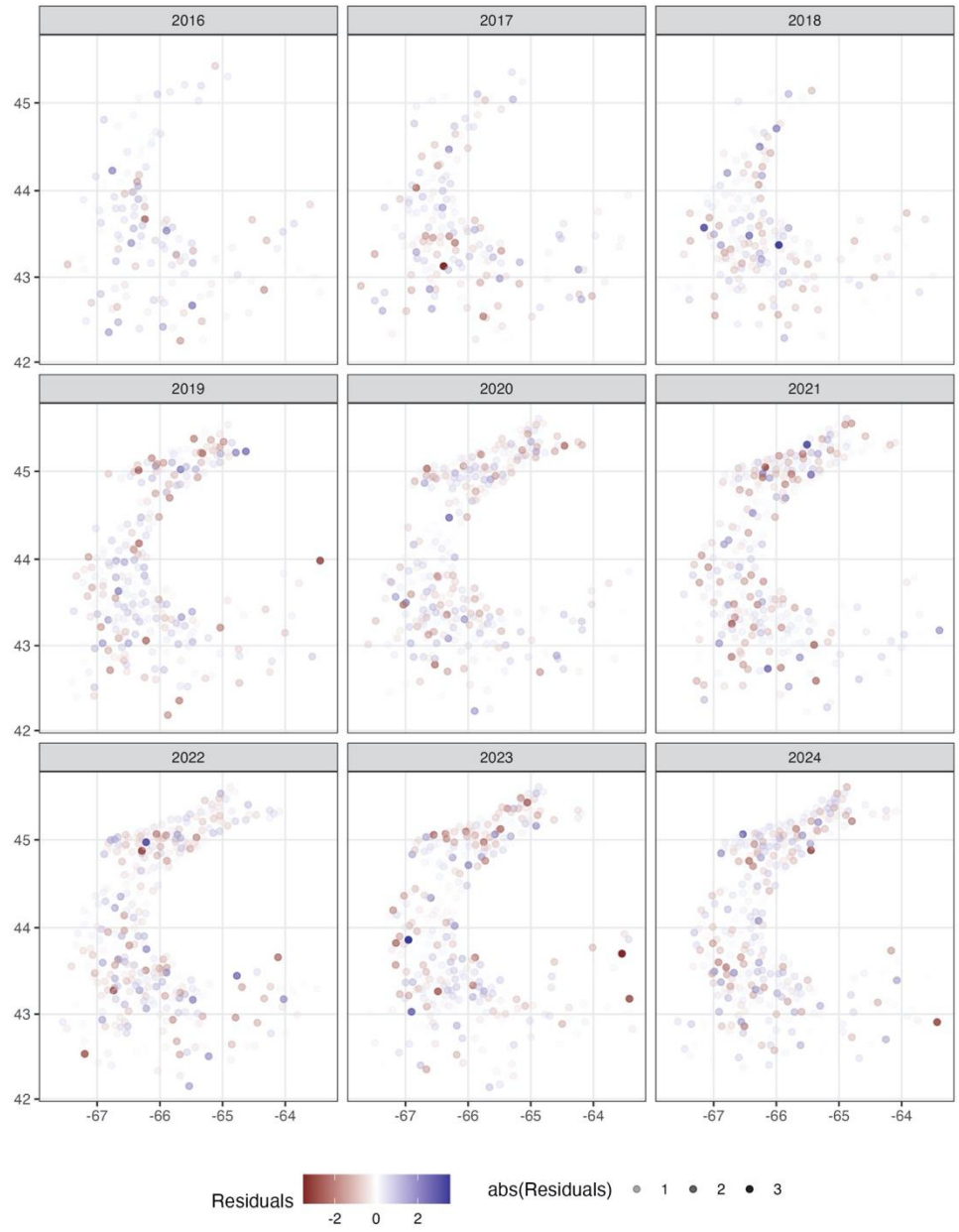


Figure 20. Estimation residuals for the cod density model. Colors indicate residual values and color intensity indicates the scale (absolute values). There were not apparent residuals patterns.

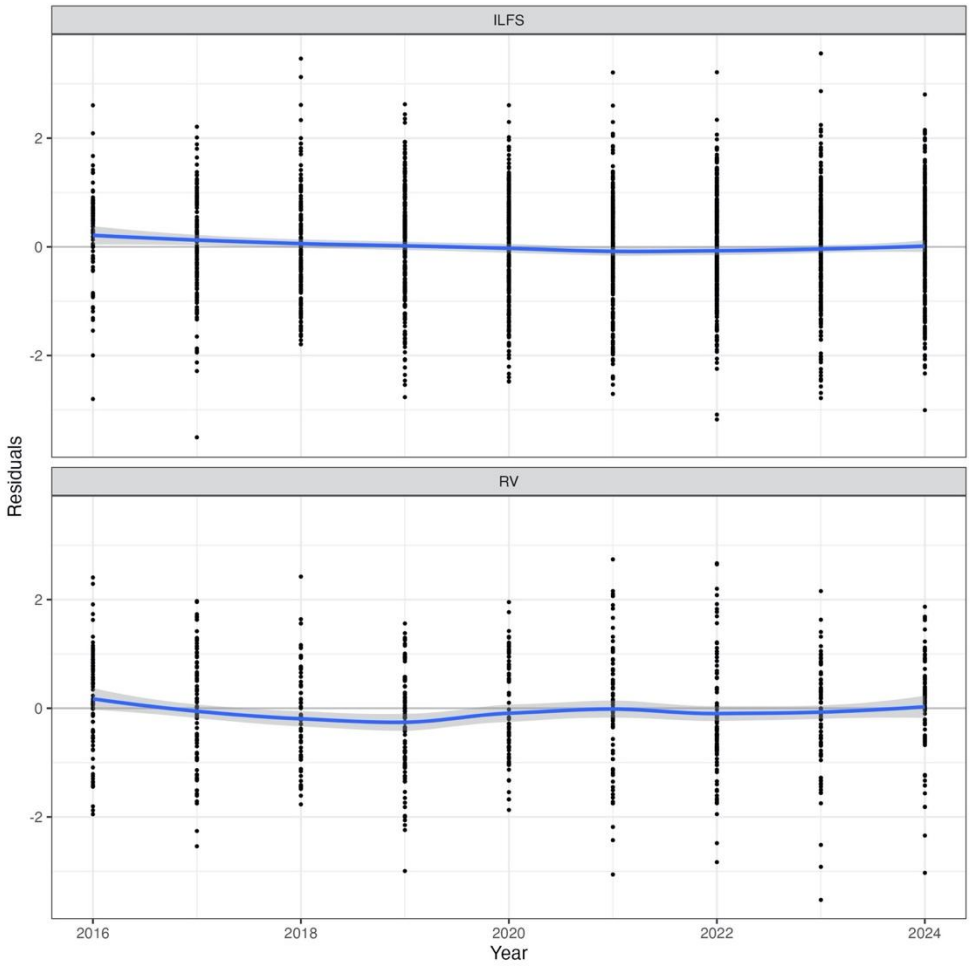


Figure 21. Estimation residuals from the cod density model were assessed against data sources. There were no apparent deviations from zero or trends, indicating good model fit.

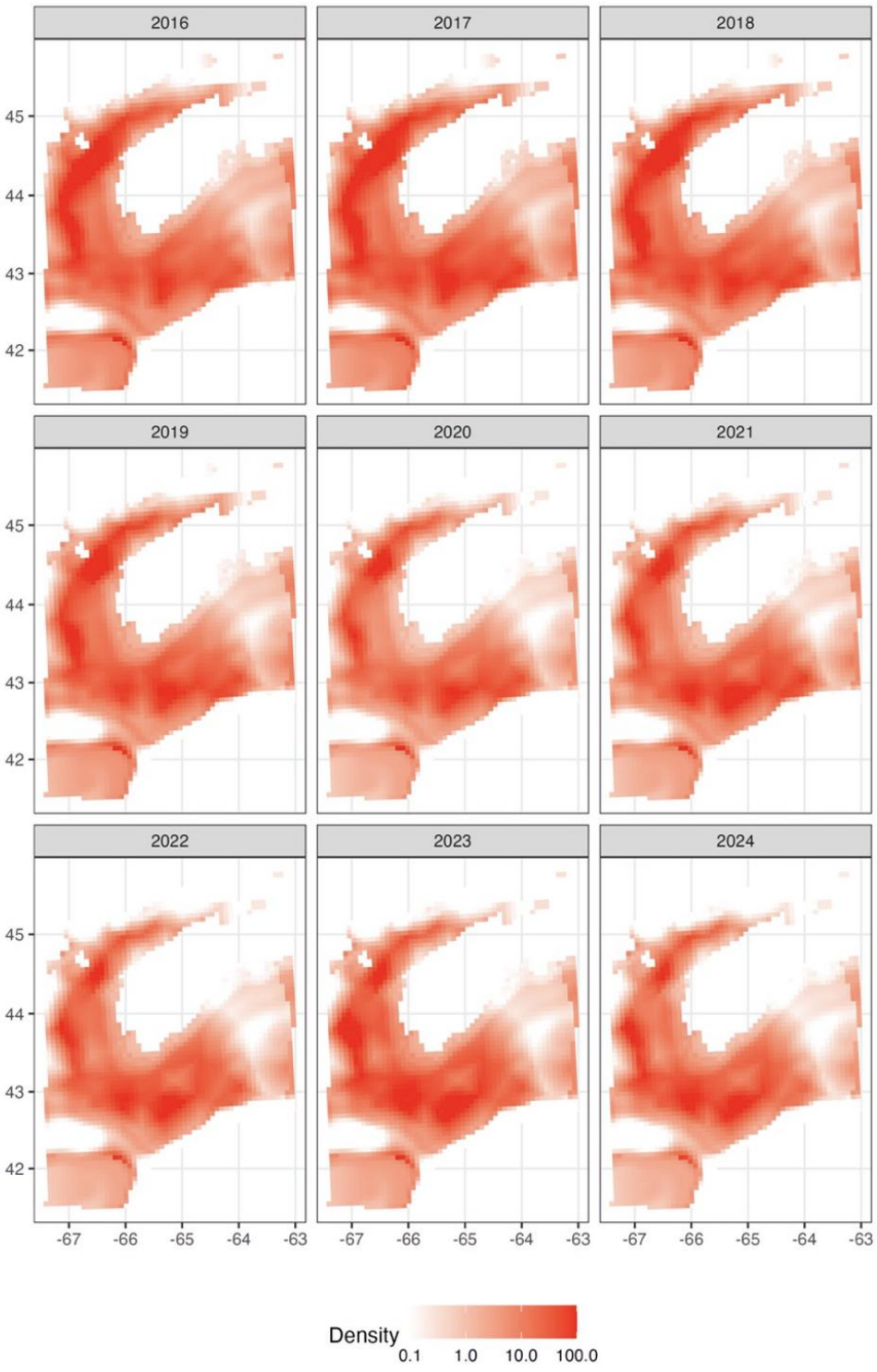


Figure 22. Estimated relative cod population density distribution.

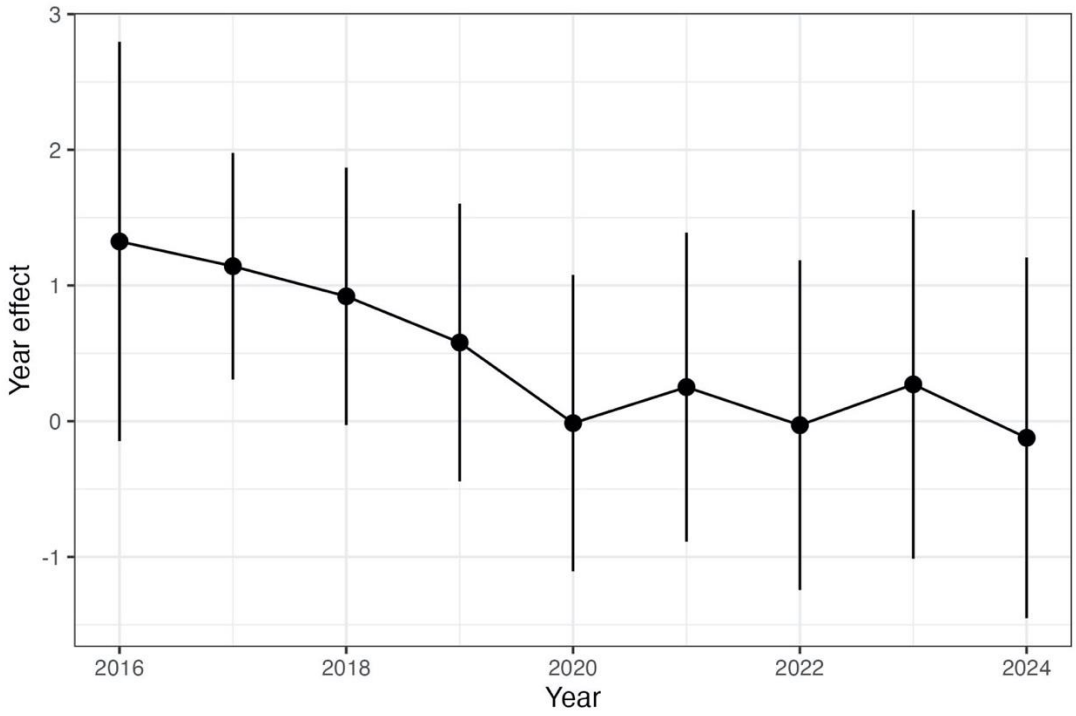


Figure 23. Estimated year effect from the cod density model indicating a slight decline in population trends during 2016-2020 and stabilization afterwards (with the log-link function, the year effect should be interpreted on the log scale).

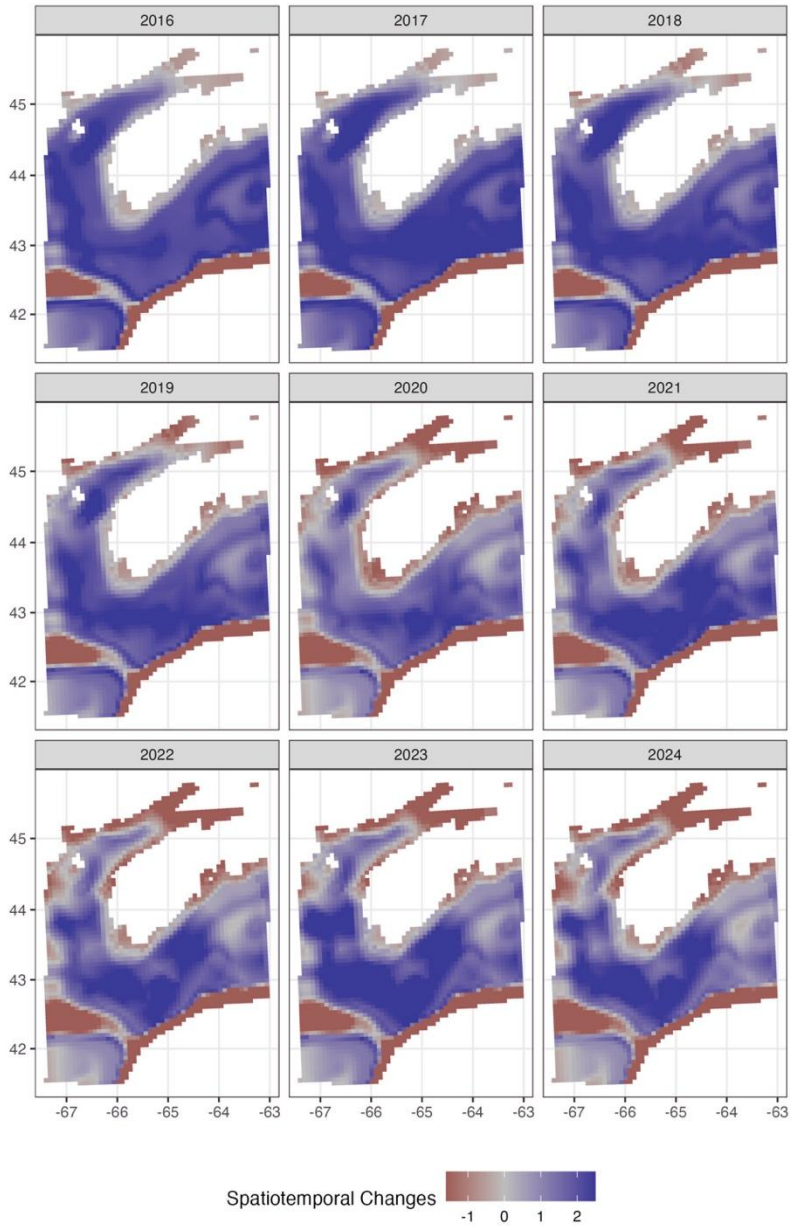


Figure 24. Estimated spatiotemporal changes during the study period (using both estimated fixed effects and the dynamic spatiotemporal effect). Color scales were set to 10%-90% quantile for visualization and grids deeper than 500m were excluded where there is no lobster fishing activity.

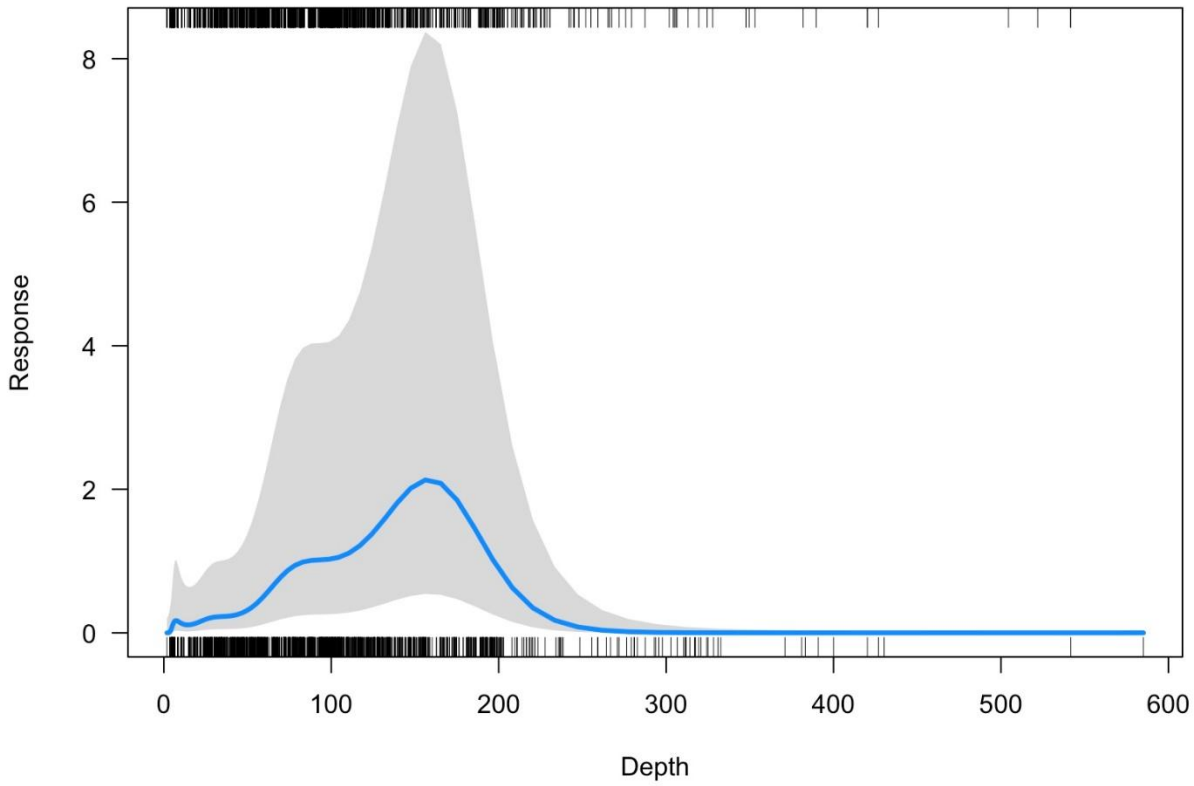


Figure 25. Model-estimated relationship between cod density and depth (blue line).

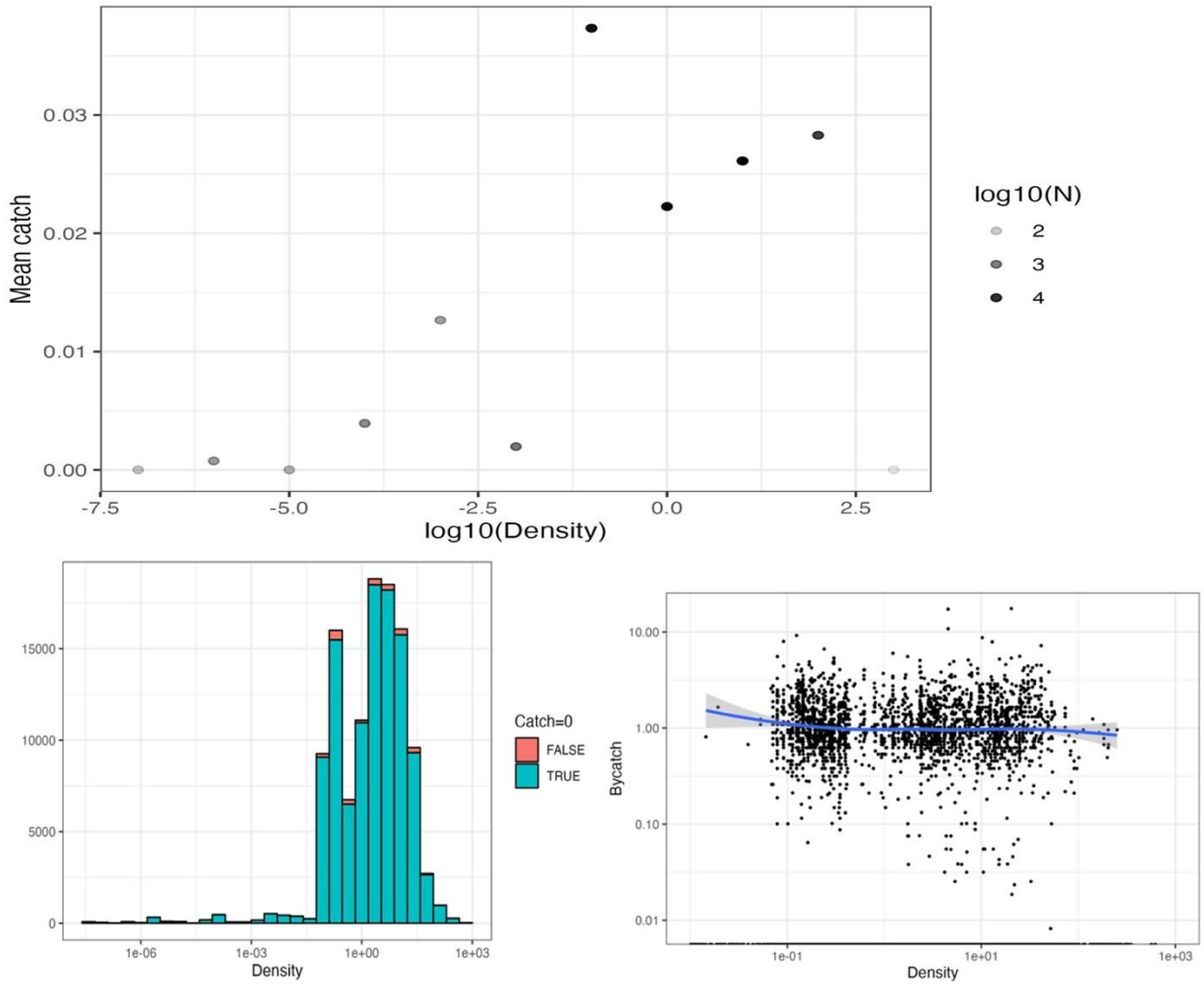


Figure 26. Preliminary analysis to assess the relationship between cod bycatch and cod density. The top panel used aggregated at-sea sampling (AS) data (based on intervals of cod density; color intensity indicates sample size, i.e., number of observations within each interval). The bottom two panels used original AS data (at the individual trap level) and plot the proportion of zeros and positive catches, respectively.

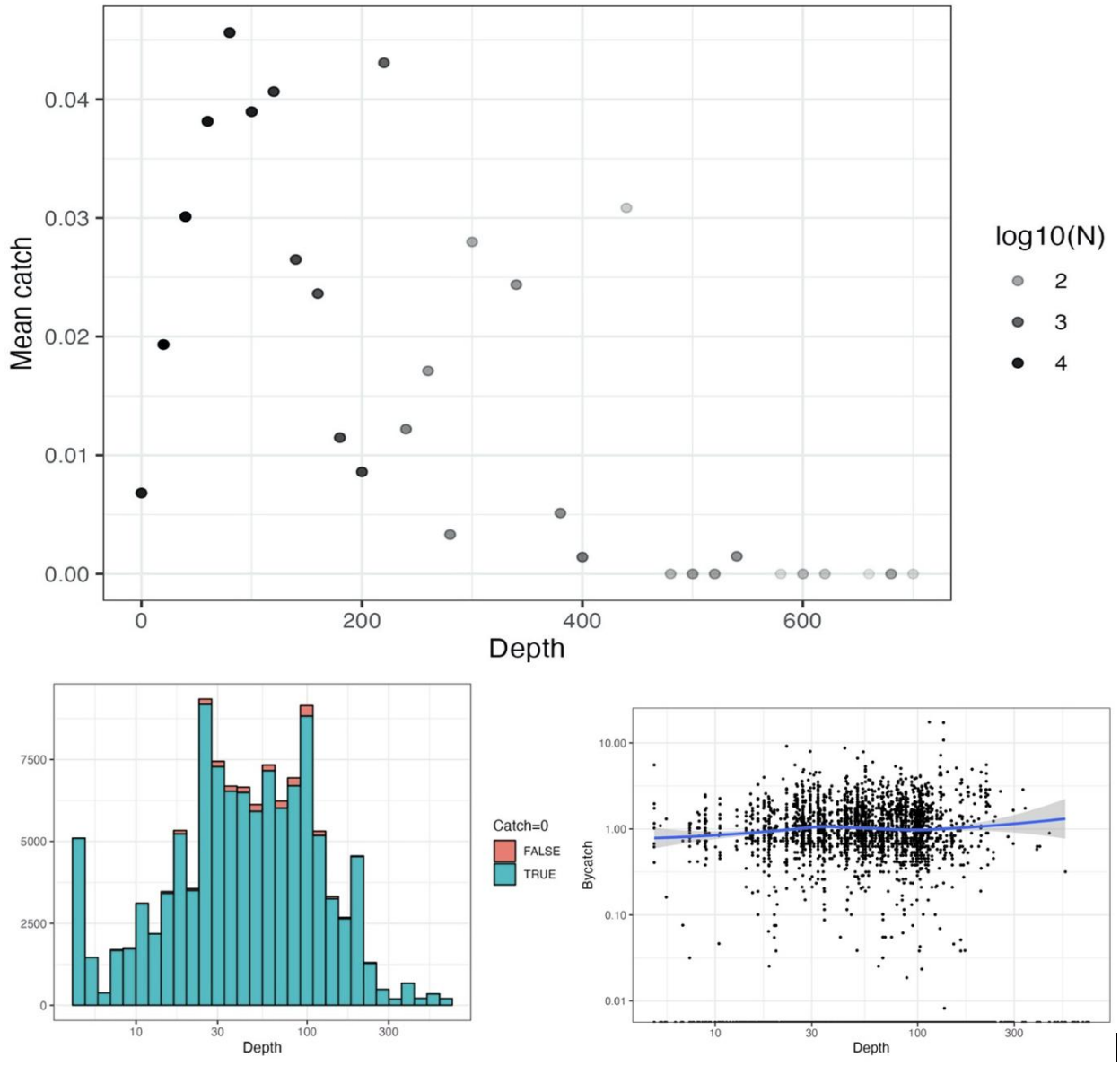


Figure 27. Preliminary analysis to assess the relationship between cod bycatch and depth. The top panel used aggregated at-sea sampling (AS) data (based on intervals of depth; color intensity indicates sample size, i.e., number of observations within each interval). The bottom two panels used original AS data (at the individual trap level) and plot the proportion of zeros and positive catches, respectively.

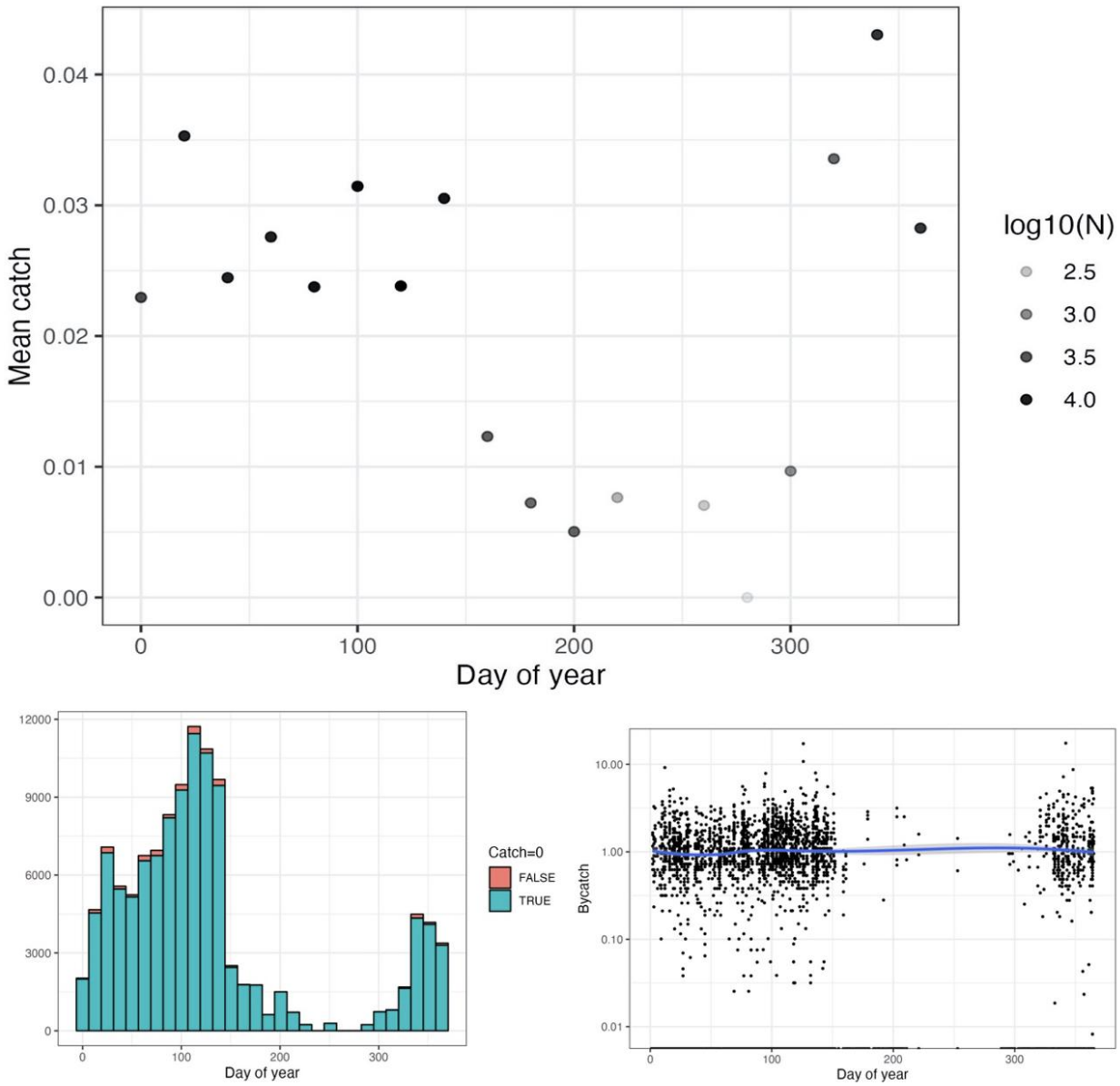


Figure 28. Preliminary analysis to assess the relationship between cod bycatch and day of year. The top panel used aggregated at-sea sampling (AS) data (based on intervals of day of year; color intensity indicates sample size, i.e., number of observations within each interval). The bottom two panels used original AS data (at the individual trap level) and plot the proportion of zeros and positive catches, respectively.

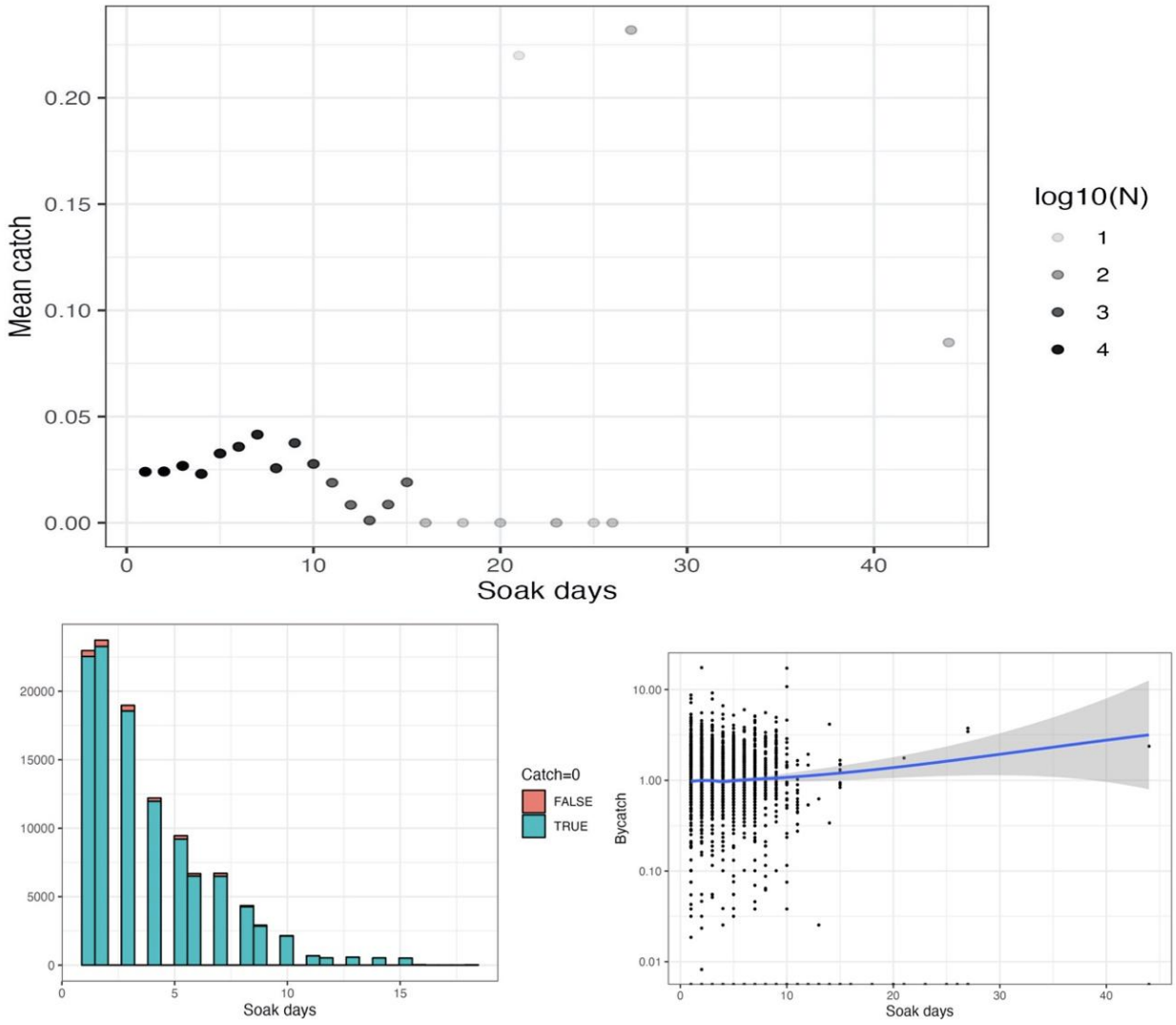


Figure 29. Preliminary analysis to assess the relationship between cod bycatch and soak days. The top panel used aggregated at-sea sampling (AS) data (based on soak days; color intensity indicates sample size, i.e., number of observations within each interval). The bottom two panels used original AS data (at the individual trap level) and plot the proportion of zeros and positive catches, respectively.

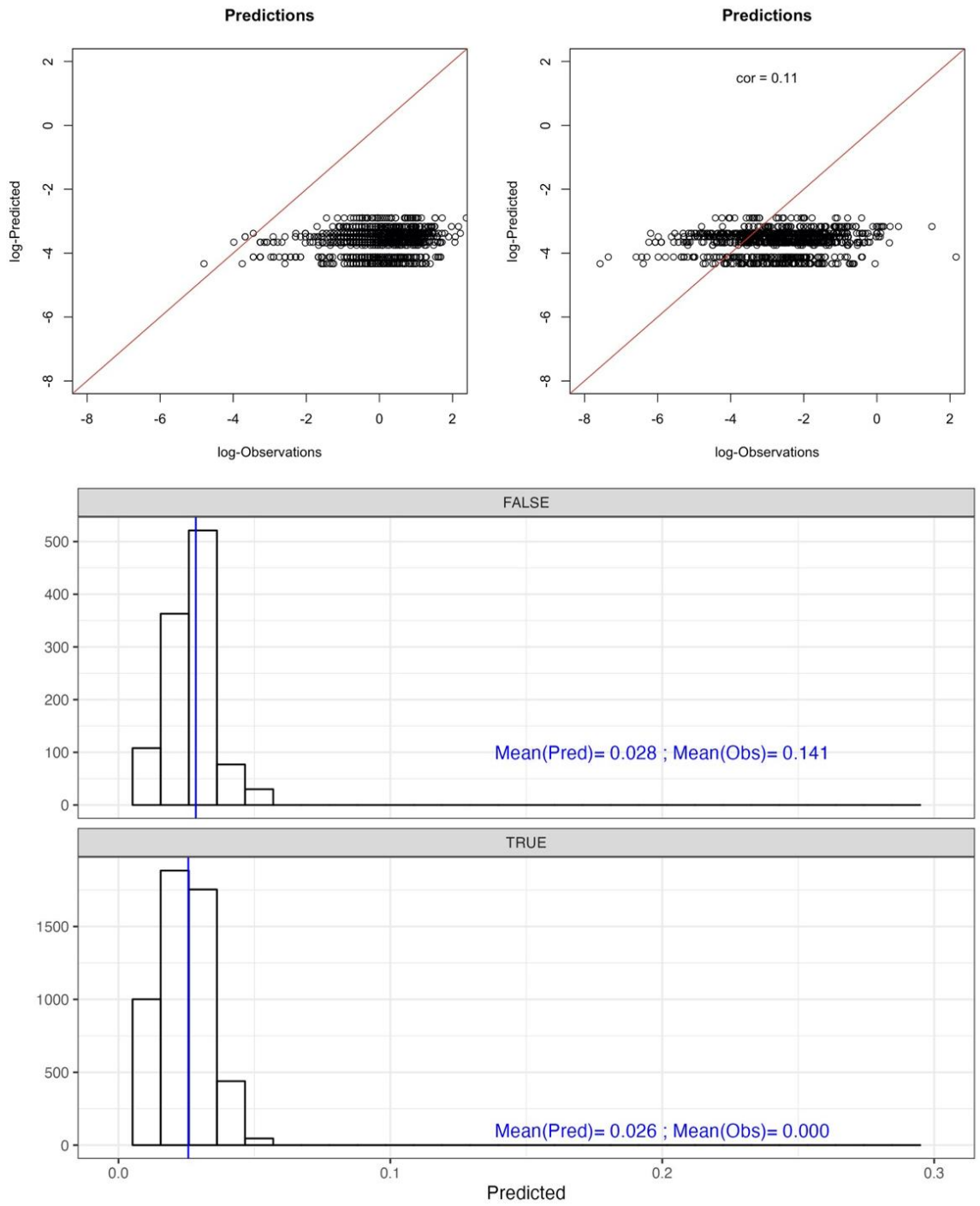


Figure 30. Residual analysis of the mean estimator (ME). Top left panel shows observations versus predictions using original data (positive values only) and top right panel show observations versus predictions aggregated to the bathymetry grid to reduce individual trap-level noise. The middle panel is residual histogram for data with positive values and the bottom panel is residual histogram for data with a zero value.

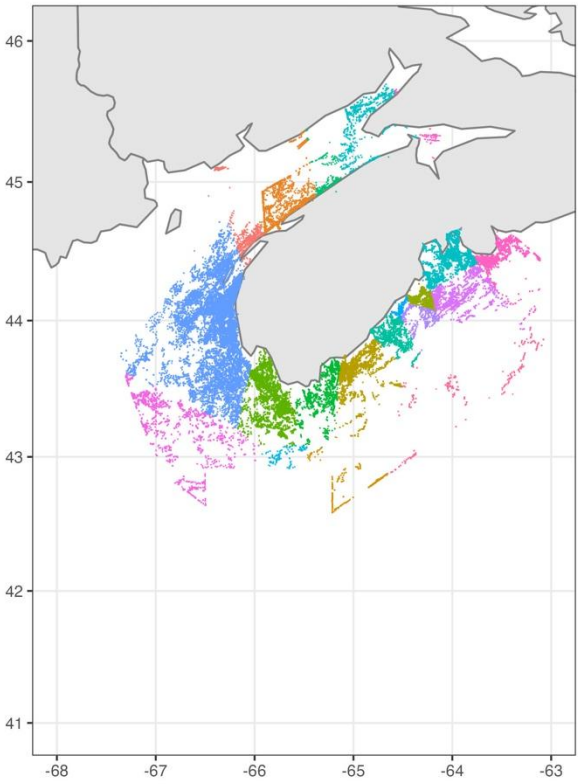


Figure 31. A post-stratification scheme tested in model development in this study, e.g., for generating a stratified mean estimator. Grouping of at-sea sampling (AS) locations was based on K-means (with 20 spatial clusters, a.k.a, "knots") and Euclidean distance (Universal Transverse Mercator coordinates) between locations. Locations and clusters associated with Lobster Fishing Area 41 were removed for privacy protection.

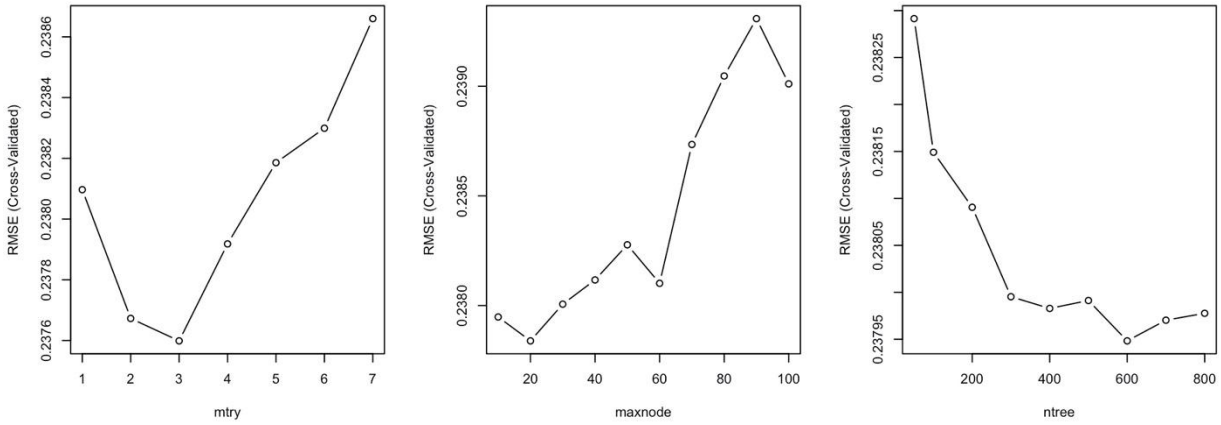


Figure 32. Results of grid-search to optimize tuning parameters for the random forest algorithm. The best parameters should reduce cross-validated root mean square errors (RMSEs); for the number of trees, the best parameter should stabilize RMSEs.

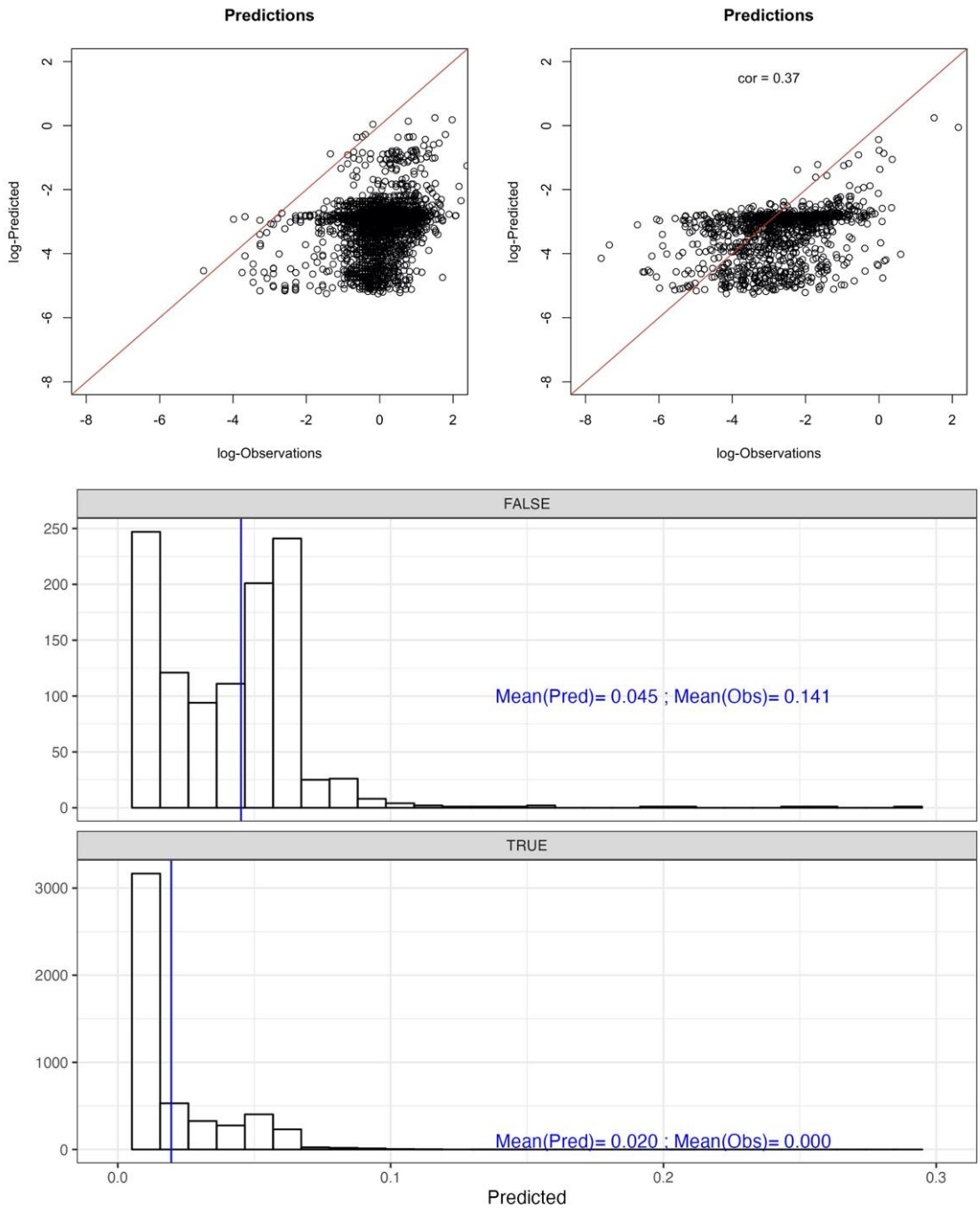
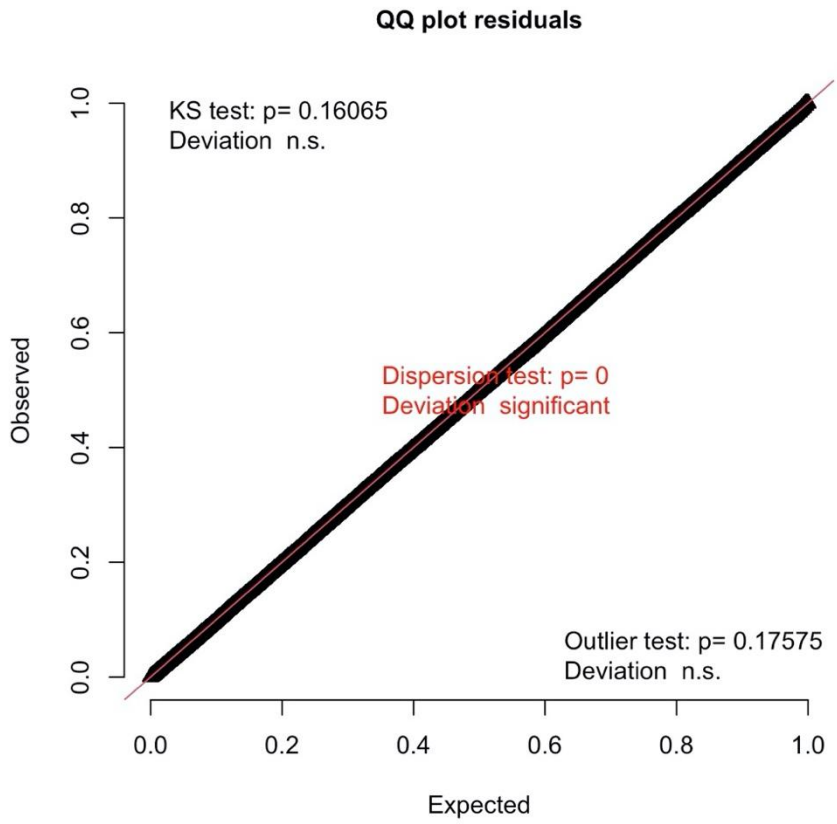


Figure 33. Residual analysis of the Random Forest (RF). Top left panel shows observations versus predictions using original data (positive values only) and top right panel show observations versus predictions aggregated to the bathymetry grid to reduce individual trap-level noise. The middle panel is residual histogram for data with positive values and the bottom panel is residual histogram for data with a zero value.



Figures 34. Residual diagnostics for the Generalized Additive Model (GAM) using the DHARMa R package indicated acceptable fit except that the parametric Tweedie distribution is likely a mis-representation of the bycatch data where over 98% are zeros.

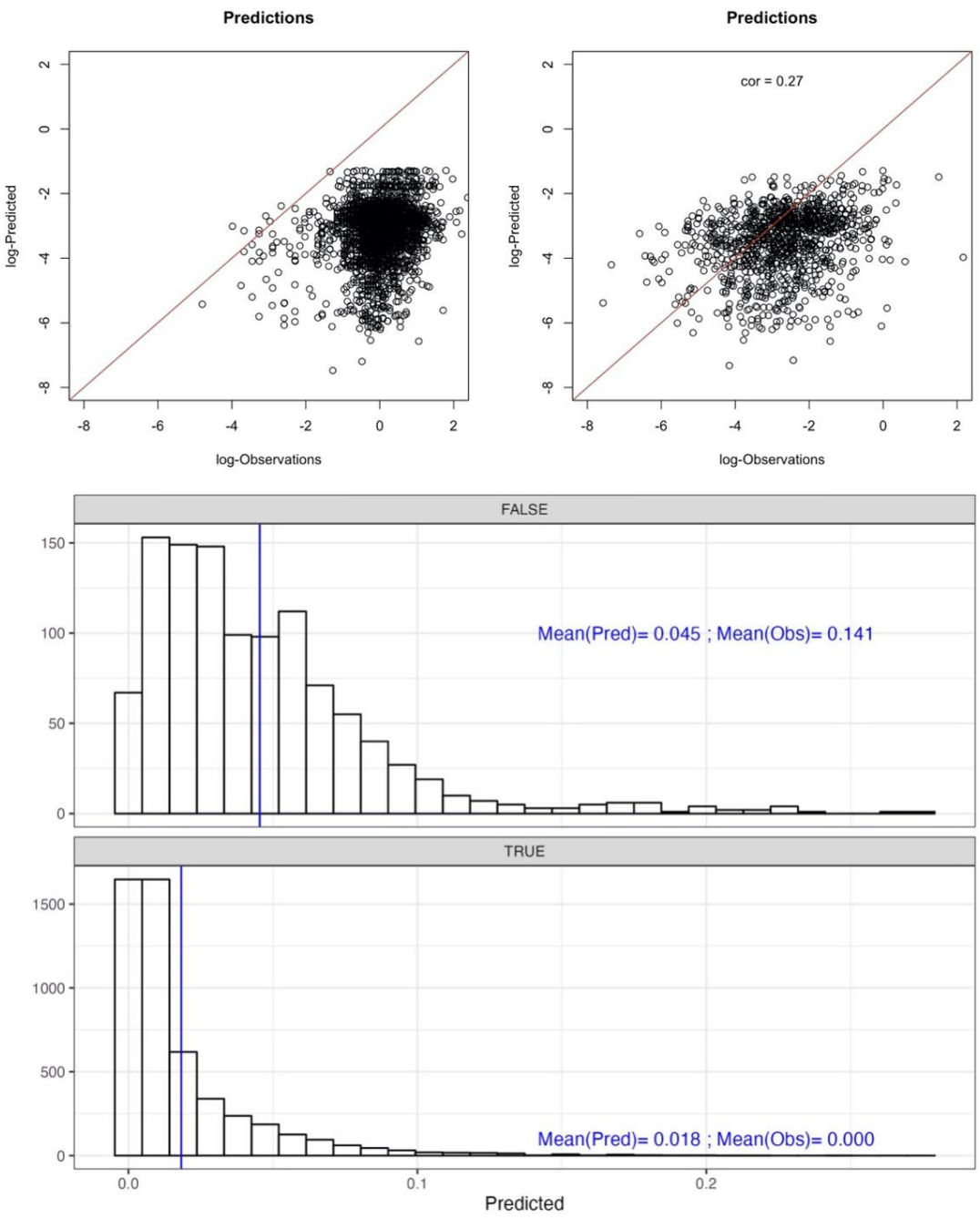


Figure 35. Residual analysis of the Generalized Additive Model (GAM). Top left panel shows observations versus predictions using original data (positive values only) and top right panel show observations versus predictions aggregated to the bathymetry grid to reduce individual trap-level noise. The middle panel is residual histogram for data with positive values and the bottom panel is residual histogram for data with a zero value.

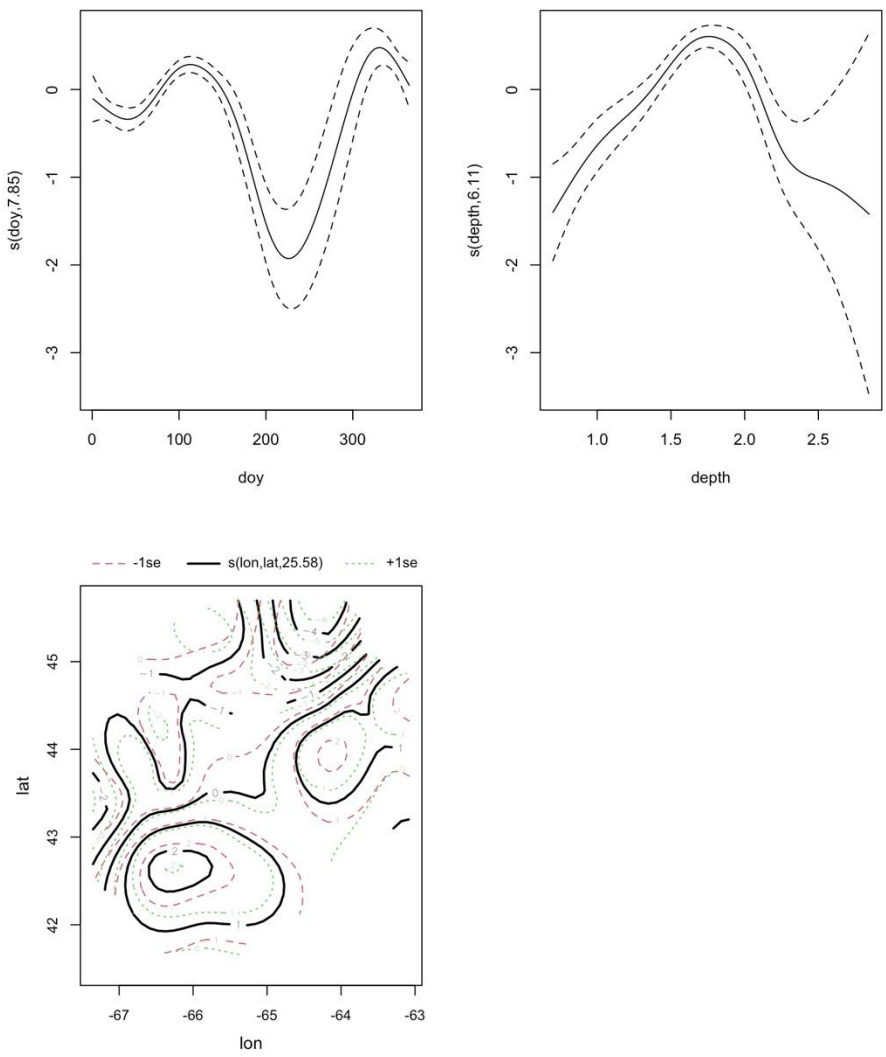


Figure 36. Generalized Additive Model (GAM) estimated relationships between cod bycatch and the three predictors, day of year (doy), depth and spatial location (longitude and latitude), respectively.

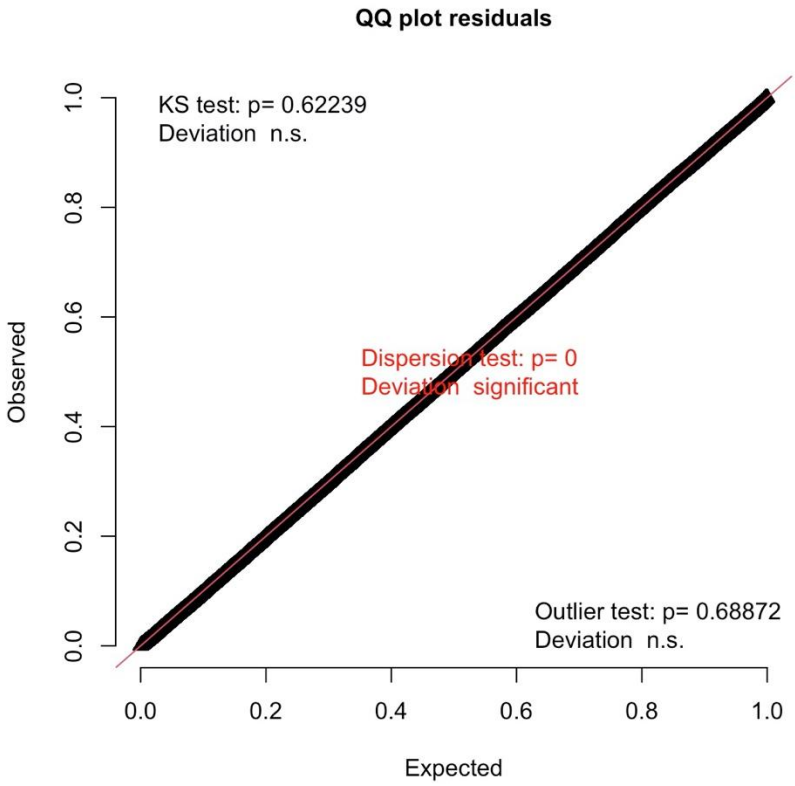


Figure 37. Residual diagnostics for the Spatio-Temporal Model (STM) using the DHARMa R package indicated acceptable fit except that the parametric Tweedie distribution is likely a mis-representation of the bycatch data where over 98% are zeros.

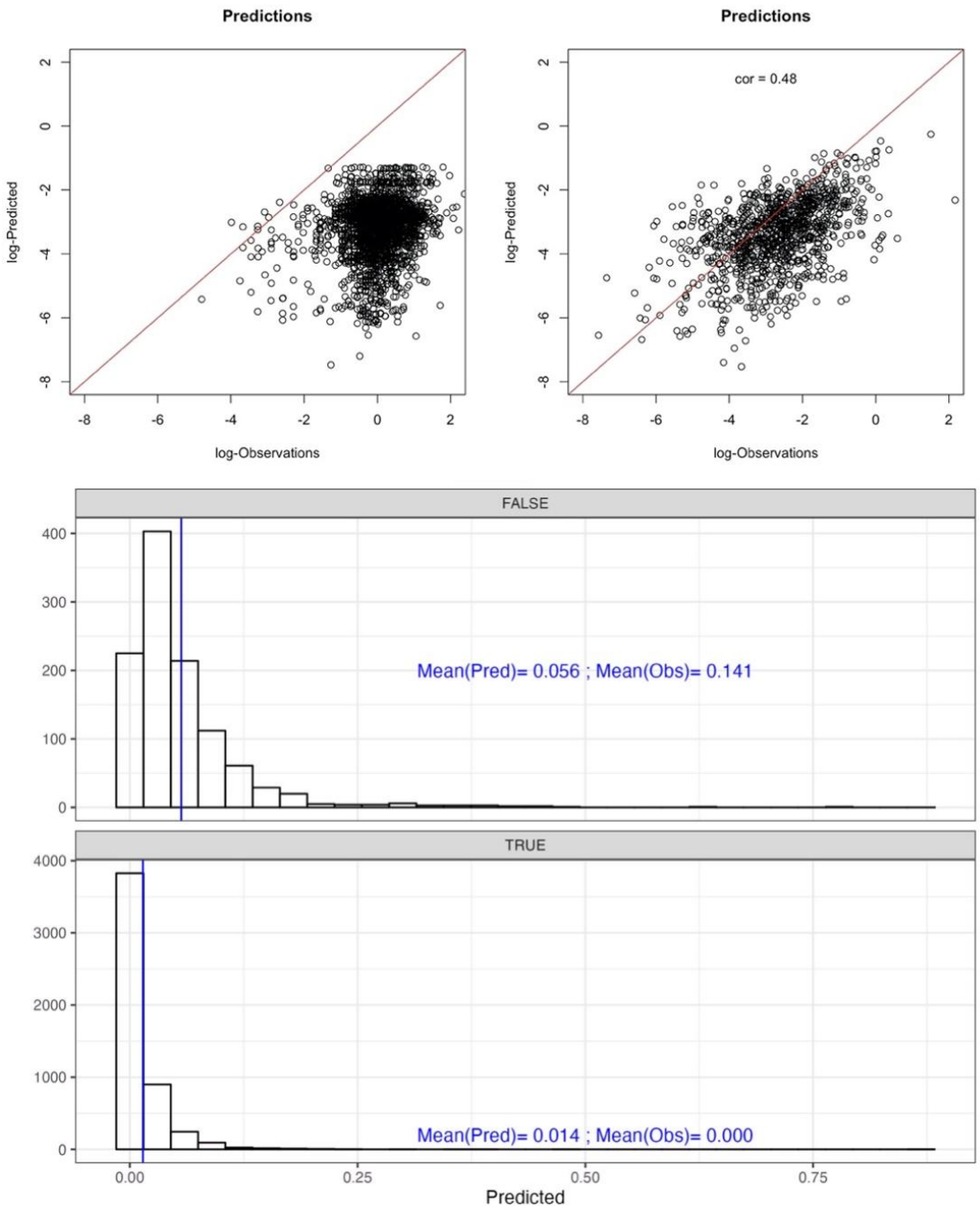


Figure 38. Residual analysis of the Spatio-Temporal Model (STM). Top left panel shows observations versus predictions using original data (positive values only) and top right panel show observations versus predictions aggregated to the bathymetry grid to reduce individual trap-level noise. The middle panel is residual histogram for data with positive values and the bottom panel is residual histogram for data with a zero value. The model tended to underpredict positive observations more than overpredict zero observations, possibly due to an extremely skewed proportion of zeros.

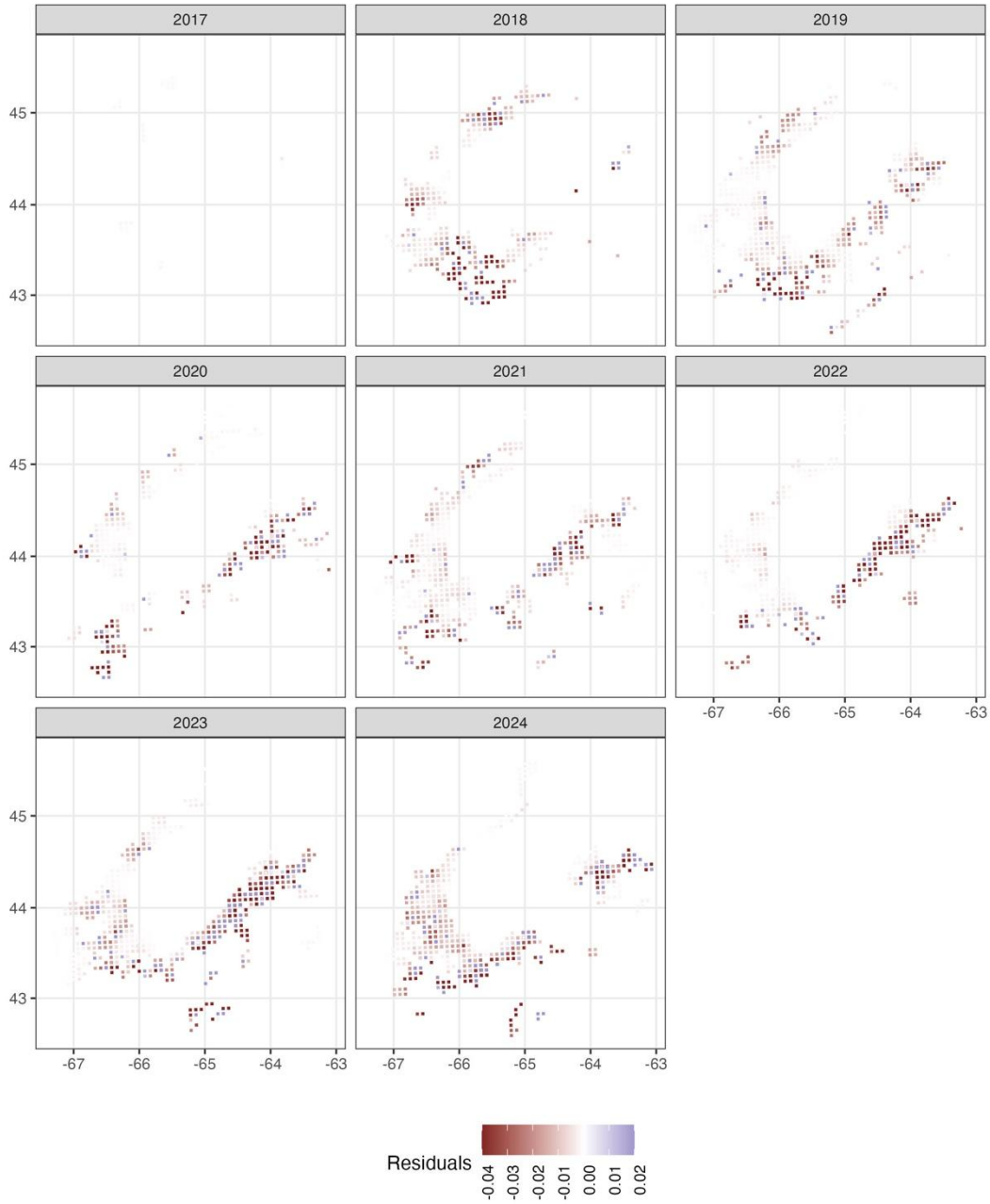


Figure 39. Estimation residuals for the cod bycatch model aggregated to the bathymetry grid. Colors indicate residual values and color intensity indicates the scale (absolute values). The model tended to underpredict positive observations more than overpredict zero observations, possibly due to a predominant proportion of zeros. Grid cells associated with lobster fishing area 41 were removed for privacy protection. All observed trips in 2016 were in LFA 41.

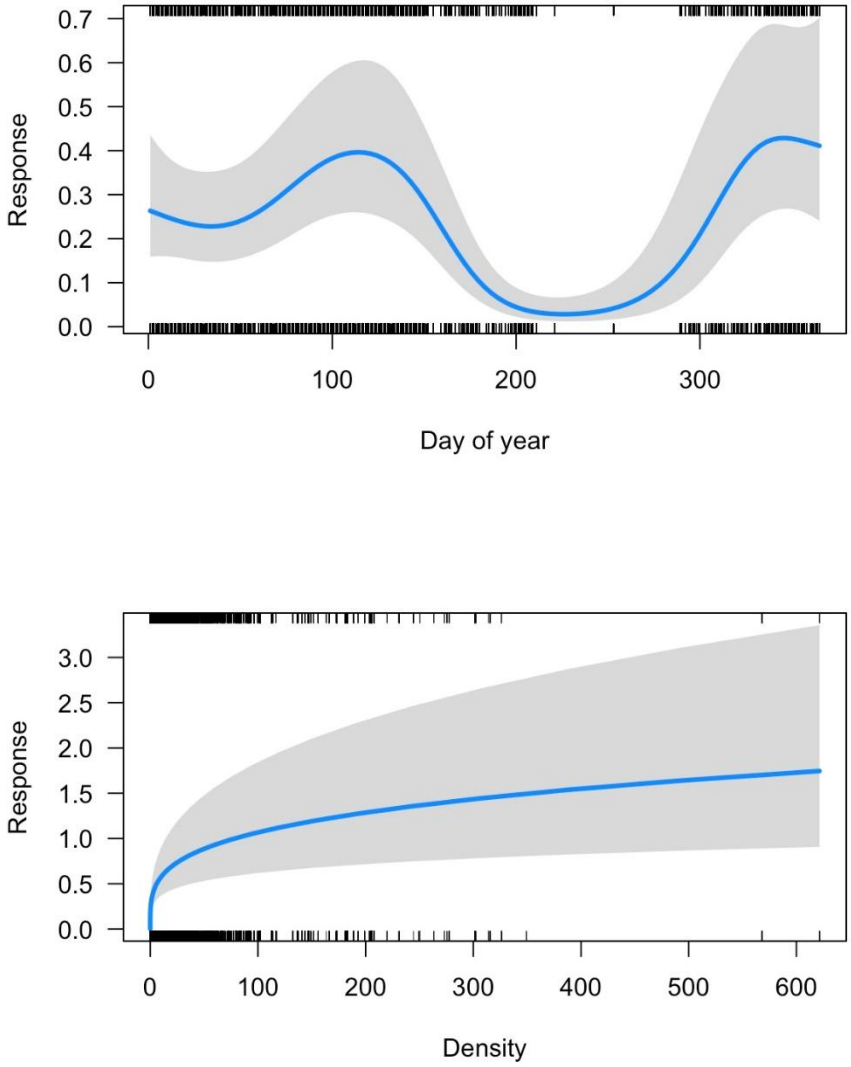


Figure 40. Spatio-Temporal Model (STM) estimated relationships between cod bycatch and predictors, day of year, depth and cod density.

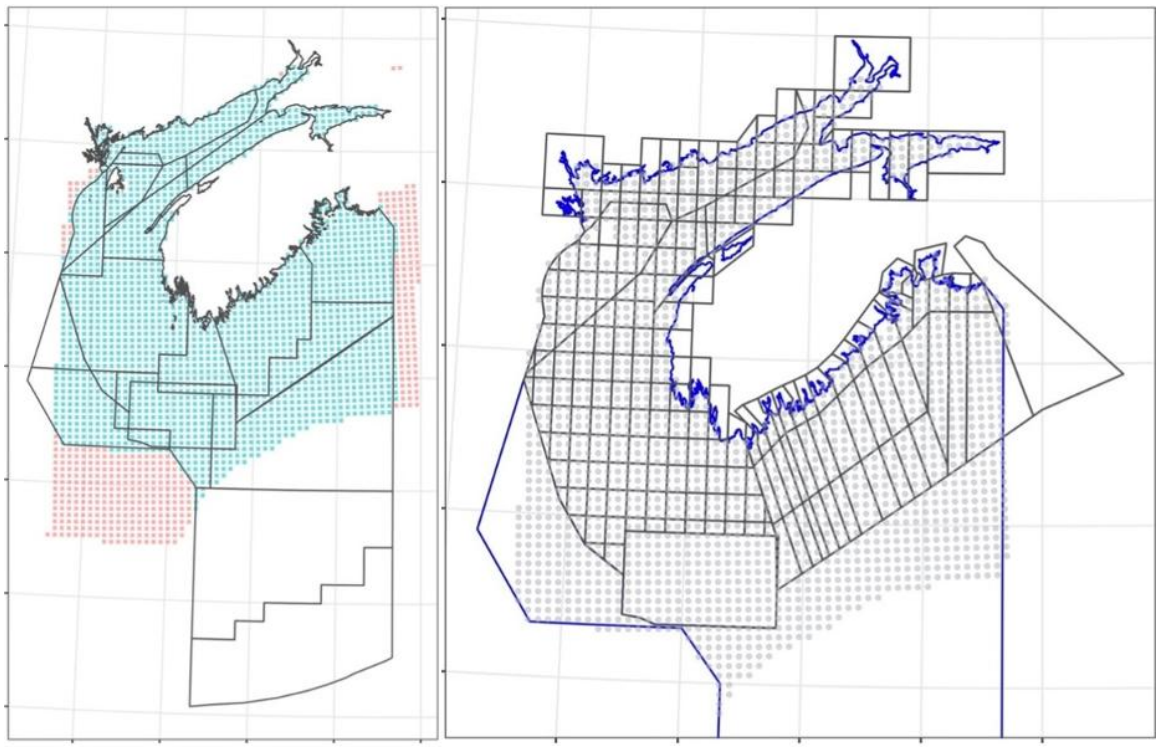


Figure 41. Relevant prediction grid for the study area were highlighted in blue on the left panel (areas with near-zero cod density or fishing effort were excluded for simplification). The right panel shows the match between the lobster grid (black polygons) and bathymetry grid (gray dots) for inshore lobster fishing areas (LFAs) and within the study area (blue lines). For LFA 41, the bathymetry grid was not used as commercial lobster fishing effort data contained information of exact locations and date.

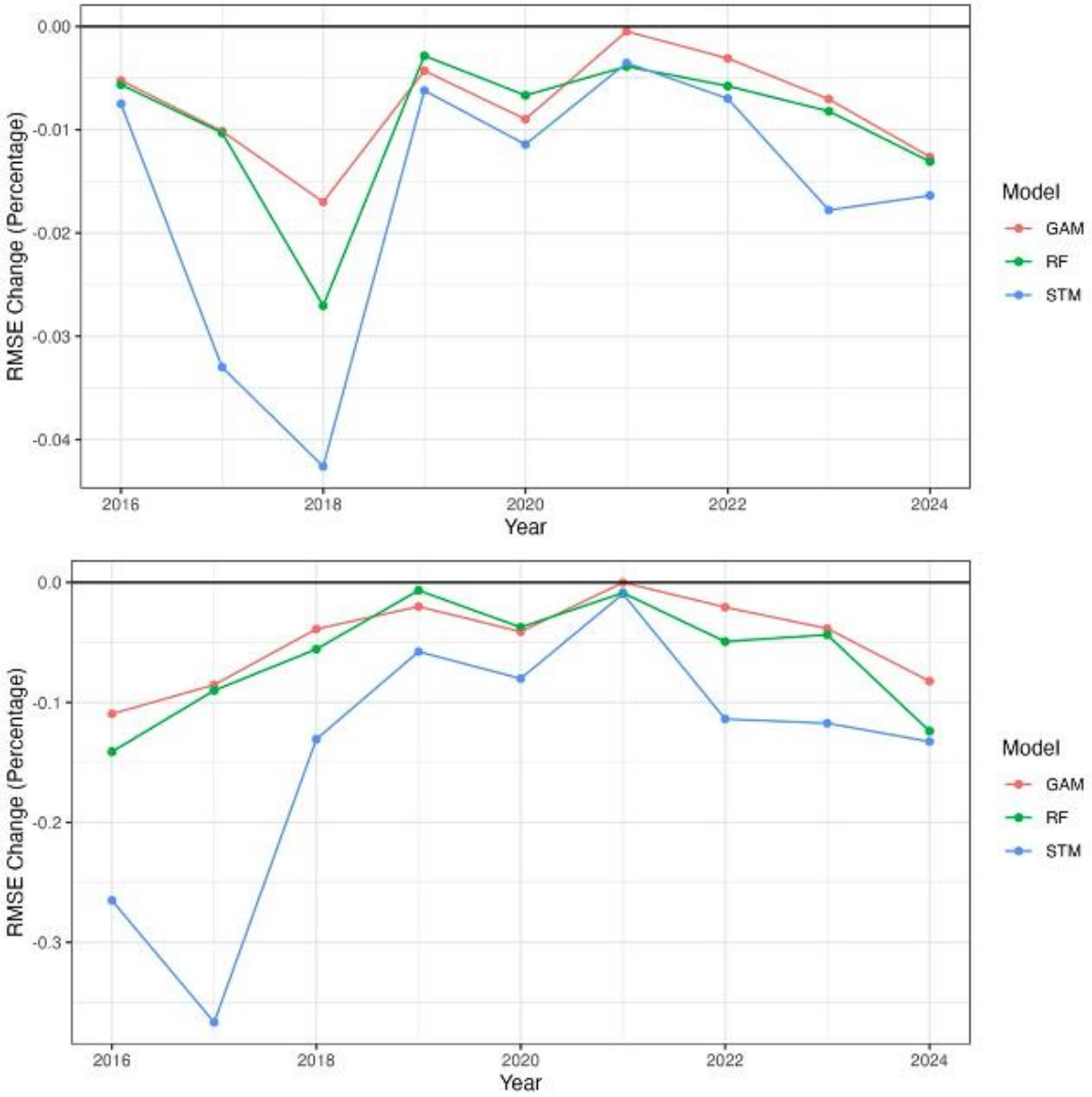


Figure 42. Reductions of cross-validated residual root mean square errors using original data resolution (top panel) and aggregated to the bathymetry grid (bottom panel), for the three model-based approaches.

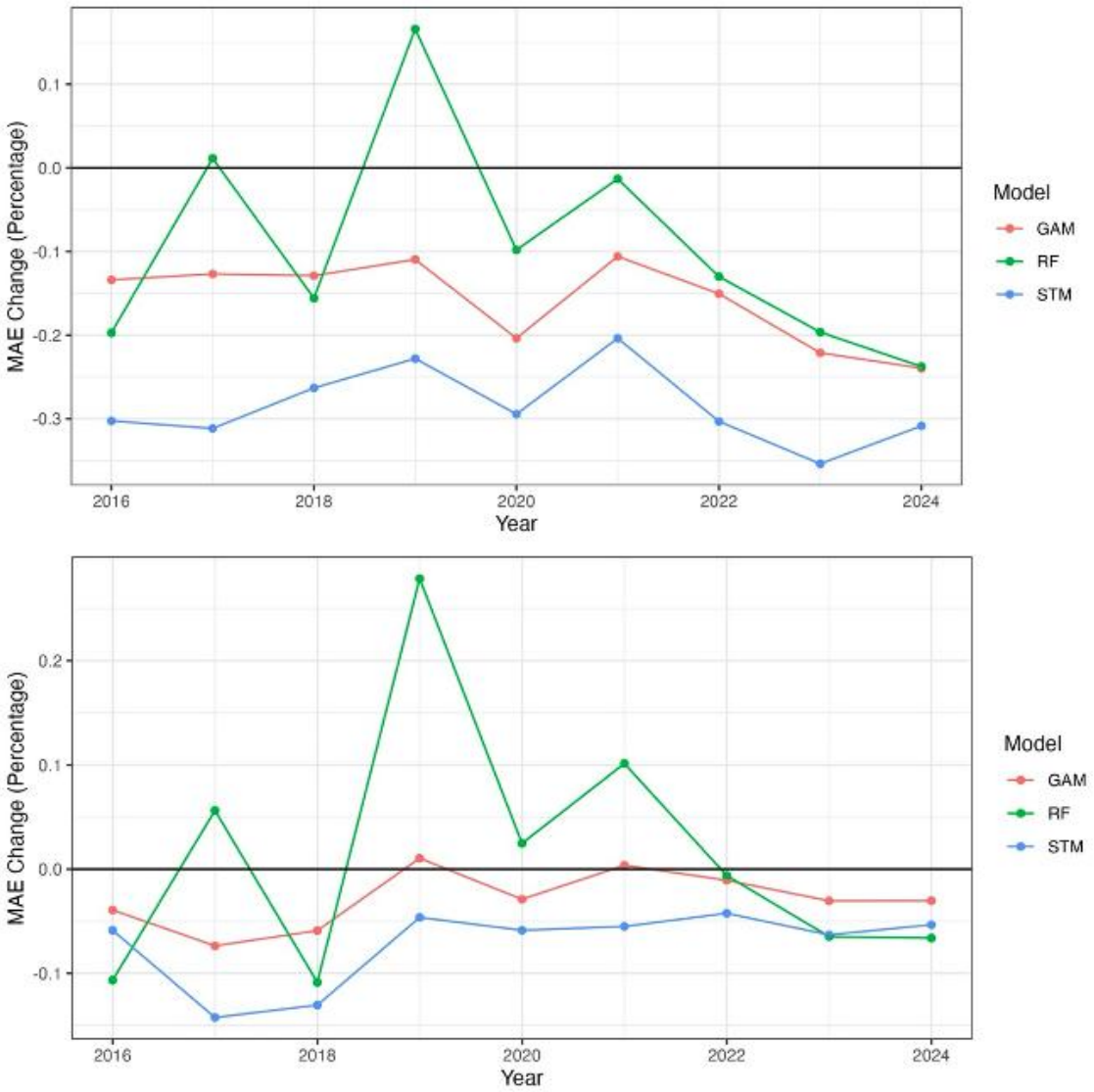


Figure 43. Reductions of cross-validated residual mean absolute errors using original data resolution (top panel) and aggregated to the bathymetry grid (bottom panel), for the three model-based approaches.

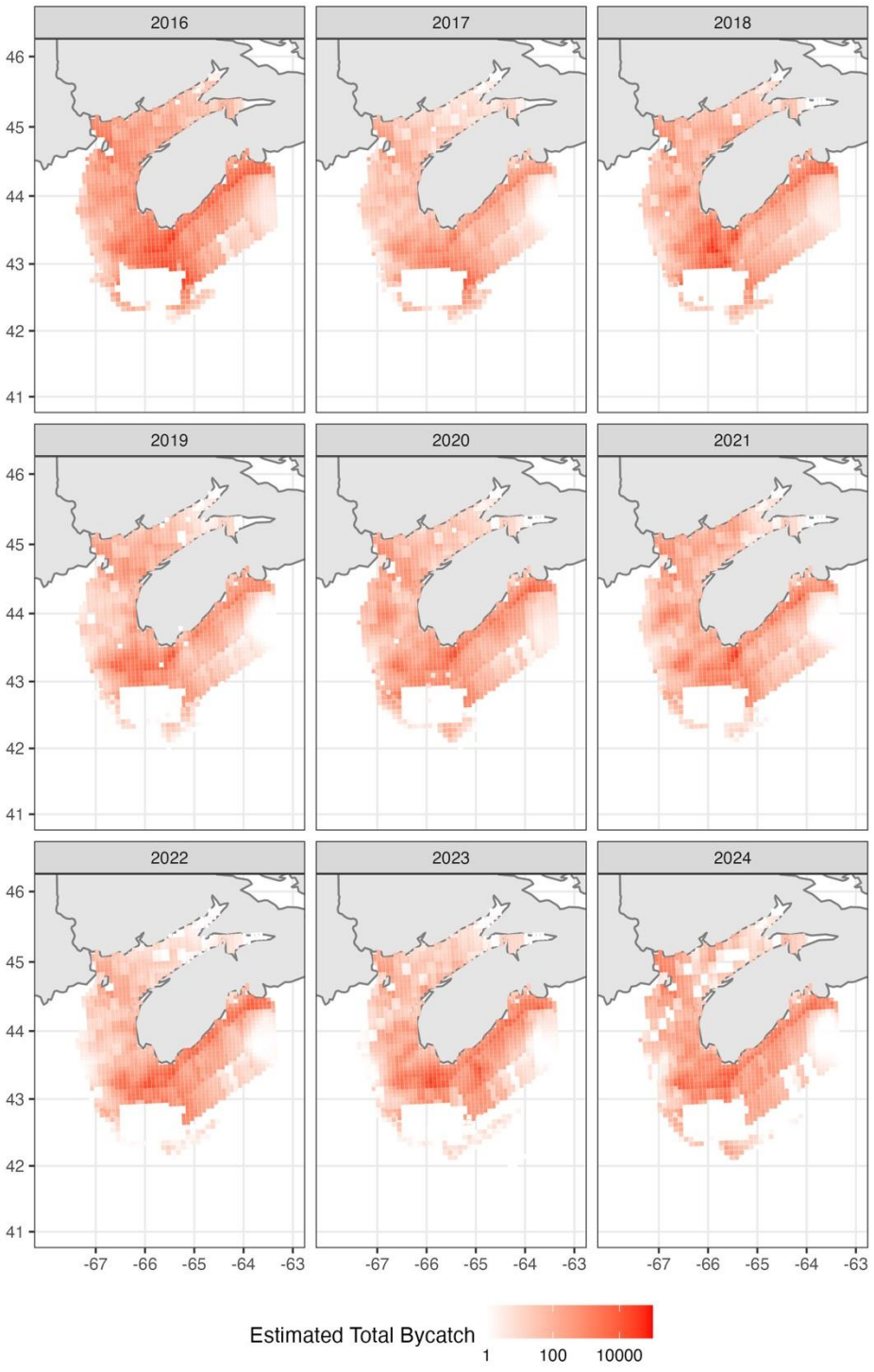


Figure 44. Spatial predictions of annual total cod bycatch in the lobster fishery in the study area (4X5Y). For lobster fishing areas (LFAs) 33-38, predictions were aggregated to the bathymetry grid; for LFA 41, predictions could be made for each fishing location but were aggregated to a 1/12 degree grid in this map for consistency of aesthetics. Predictions for 2016-2017 were not recommended for considerations given data limitations and high uncertainty.

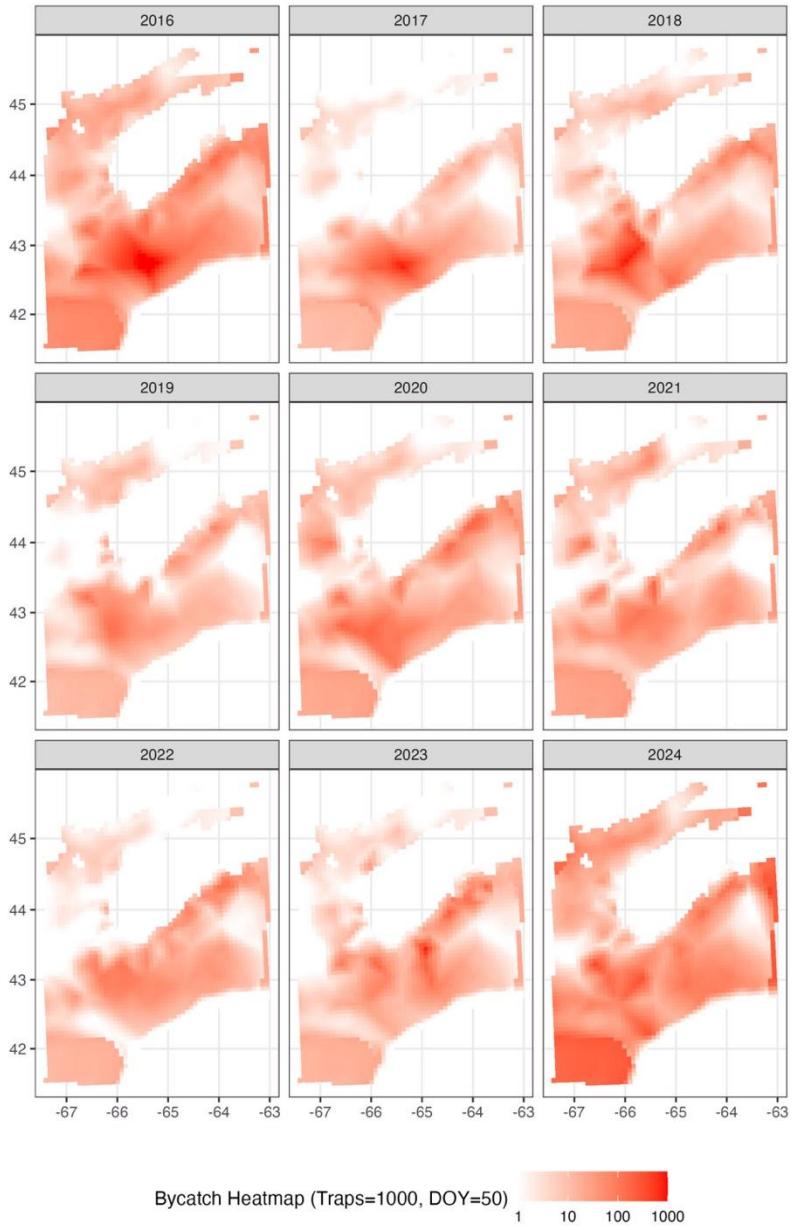


Figure 45. Spatial predictions of cod bycatch in the lobster fishery in the study area (4X5Y) assuming 1000 lobster traps within each grid (the bathymetry grid) and fished on the 50th day of year (typically high-effort time of year). Areas deeper than 500m were excluded. Predictions for 2016-2017 were not recommended for considerations given data limitations and high uncertainty.

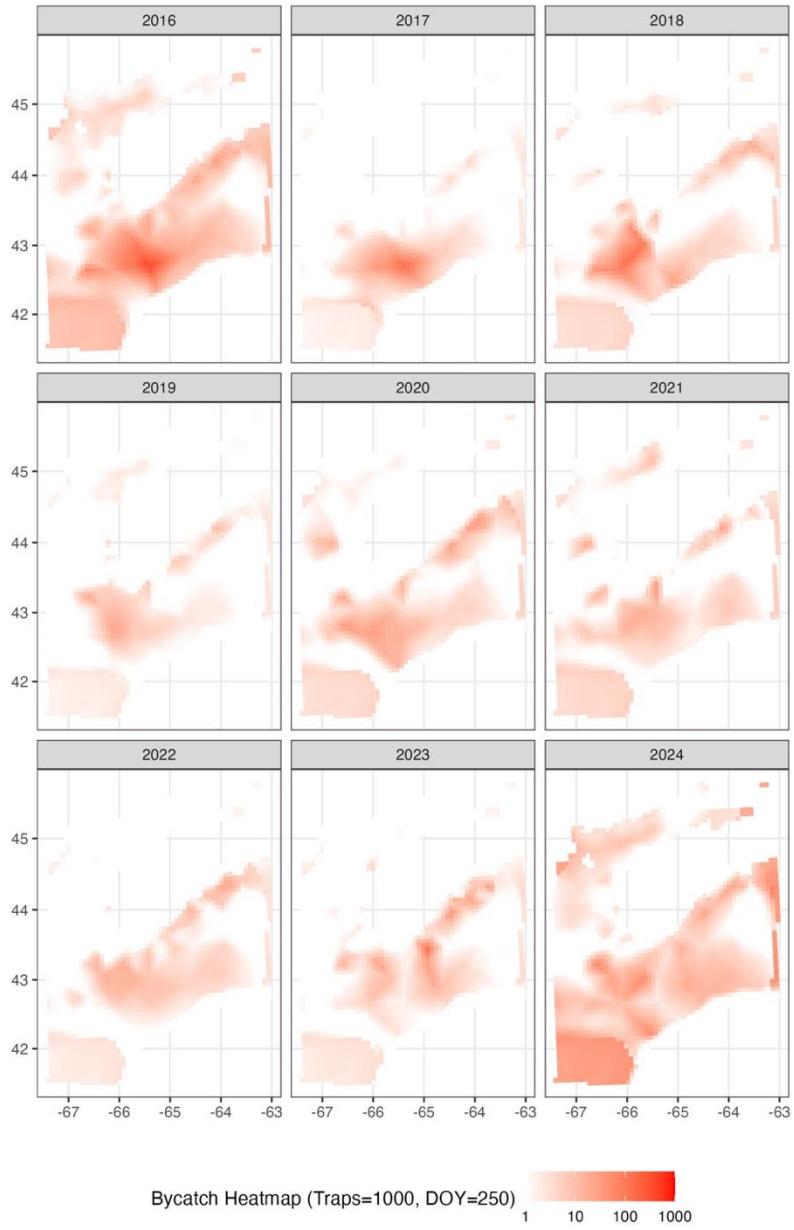


Figure 46. Spatial predictions of cod bycatch in the lobster fishery in the study area (4X5Y) assuming 1000 lobster traps within each grid (the bathymetry grid) and fished on the 250th day of year (typically low-effort time of year). Areas deeper than 500m were excluded. Predictions for 2016-2017 were not recommended for considerations given data limitations and high uncertainty.

TABLES

Table 1. Summary of data from the Maritimes Summer Ecosystem Research Vessel Survey.

Year	Mean catch	Median catch	Number of zeros	Number of tows	Percent of zeros	Standard deviation	1st quantile	3rd quantile
2016	43.68	0.03	36	84	0.43	117.74	0	20.92
2017	26.94	0	46	86	0.53	62.94	0	26.33
2018	30.88	0	49	69	0.71	100.19	0	5.93
2019	12.1	0	61	86	0.71	33.26	0	0.02
2020	15.7	0	47	71	0.66	37.07	0	6.24
2021	42.48	0	44	72	0.61	92.63	0	45.58
2022	13.47	0	71	91	0.78	41.8	0	0
2023	26.74	0	60	80	0.75	70.31	0	1.08
2024	17.22	0	44	79	0.56	40.41	0	10.44

Table 2. Summary of data for the Inshore Lobster Trawl Survey.

Year	Mean catch	Median catch	Number of zeros	Number of tows	Percent of zeros	Standard deviation	1st quantile	3rd quantile
2016	6.29	1.24	32	93	0.34	12.46	0.00	7.75
2017	15.08	0.00	70	128	0.55	40.58	0.00	10.60
2018	14.85	0.00	92	149	0.62	40.53	0.00	6.77
2019	9.91	0.00	138	240	0.58	26.25	0.00	7.38
2020	4.34	0.00	197	267	0.74	16.56	0.00	0.73
2021	3.63	0.00	221	288	0.77	11.45	0.00	0.00
2022	5.34	0.00	235	308	0.76	16.51	0.00	0.00
2023	11.75	0.00	197	288	0.68	95.09	0.00	3.17
2024	5.65	0.00	213	293	0.73	16.12	0.00	0.67

Table 3. Summary of at-sea observer data used in this study from lobster fishing areas 33–41.

Year	Total number of observed traps	Number of positive cod bycatch	Total cod bycatch (kg)	Median catch	Mean catch	Standard deviation	Coefficient of variation	Sampling Ratio
2016	2,037	58	111.97	0	0.055	0.5439	9.89	0.0061%
2017	2,400	36	56.19	0	0.0234	0.2258	9.64	0.0078%
2018	5,359	141	225.76	0	0.0421	0.3317	7.87	0.0157%
2019	14,413	159	189.58	0	0.0132	0.1465	11.14	0.0438%
2020	7,400	168	187.3	0	0.0253	0.1977	7.81	0.0254%
2021	16,254	220	263.83	0	0.0162	0.2172	13.38	0.0505%
2022	13,242	273	342.01	0	0.0258	0.2222	8.6	0.0428%
2023	24,839	658	759.76	0	0.0306	0.2256	7.38	0.0814%
2024	27,347	717	930.08	0	0.034	0.2491	7.32	0.1475%
2016-2024	113,291	2,430	3066.47	0	0.0271	0.2359	8.71	0.0416%

Table 4. Summary of commercial lobster fishing effort data (trap hauls rounded to the nearest integers).

Year	Total effort within LFAs 33–41	Total effort within 4X5Y
2016	33,404,188	32,584,884
2017	30,654,829	29,891,246
2018	34,121,958	33,336,893
2019	32,927,578	32,257,729
2020	29,122,810	28,482,123
2021	32,214,057	31,525,358
2022	30,933,528	30,271,136
2023	30,525,150	29,841,533
2024	18,539,563	18,139,940

*Table 5. Length-weight relationship model parameters used for conversion from catch numbers and length frequency to catch weights (Andrushchenko et al. 2018; parameters alpha and beta correspond to the parameters in the conversion model weight=alpha*length^beta).*

Quarter	alpha	beta
Q1	0.0000081	3.0503
Q2	0.0000084	3.0410
Q3	0.0000087	3.0233
Q4	0.0000063	3.1152

Table 6. A list of tested models for cod density estimation and model selection. The spatial and spatiotemporal effects can be either turned "on" or "off" and the spatiotemporal effect has various assumptions of temporal autocorrelation, AR1 (autoregressive), RW (random walk) and IID (independently and identically distributed). Details of model specifications and assumptions can be found in Anderson et al. 2022. AIC=Akaike information criteria; BIC=Bayesian information criteria.

Model #	Spatial Effect	Spatiotemporal Effect	AIC	BIC
1	on	AR1	10801.9	10908.59
2	on	RW	10799.6	10900.37
3	on	IID	10806.97	10907.74
4	on	off	10816.63	10911.47
5	off	RW	10834.99	10929.82

Table 7. Summary and description of predictors and data availability in the at-sea observer (ASO) and commercial lobster fishing effort (CLFE) datasets.

Variable Type	Variable Name	ASO	CLFE	Comment
Trip	Trip, string, trap identifiers	Recorded (some missing data)	Not available	Not used as majority of CLFE data were not observed.
Trip	location	Recorded; longitude and latitude	Available on a coarser grid (the lobster grid) for inshore LFAs	Can be aggregated to various resolutions such as NAFO regions, LFAs, lobster grids, bathymetry grids or via post-stratification. Modeled using exact locations.

Variable Type	Variable Name	ASO	CLFE	Comment
Trip	time	Recorded; exact date and time	Available on a coarser grid (week of lobster fishing season) for inshore LFAs	Can be aggregated to various resolutions such as week, month etc. Modeled using day of year as a continuous variable and year as a categorical variable.
Fishing activity	Fishing related identifiers, e.g., vessel identifier, license etc.	Recorded	Available	Not used as majority of CLFE data were not observed.
Fishing activity	Soak days	Recorded	Not available	Could be estimated based on time and space.
Environmental conditions	Depth, bottom type, etc.	Depth is recorded	Not available	Given exact space and time information, many environmental variables could be derived from external data sources or analyses. Bathymetry and model-estimated cod density distributions were introduced to this study.
Catch	Composition of catch within a trap	Catch of all species was recorded	Not available	Could be effective predictors to explain individual trap-level variability but without detailed data in the CLFE data, this study did not incorporate these data.

Table 8. A list of tested models within the generalized linear model (GLM) framework for cod bycatch analysis and model selection (f.year is a categorial year effect, doy is day of year, sc.doy is the scaled version, knot is a categorical post-stratification, lon is longitude, lat is latitude, dens is cod density). GAM-generalized additive model; GLMM-generalized linear mixed model; AIC-Akaike information criteria; BIC-Bayesian information criteria.

Model #	Model Type	Model Formulation	AIC	BIC	Comment
1	GLM	f.year + sc.doy + f.knot + lon + lat + depth + dens	27205.9	27543.2	including various candidate predictors
2	GAM	f.year + s(doy) + s(dens) + s(depth) + s(lon, lat)	26816.17	27331.67	including most predictors and with smoothers

Model #	Model Type	Model Formulation	AIC	BIC	Comment
3	GAM	f.year + s(doy) + dens + s(depth) + f.knot,	27142.87	27605.47	Aggregate space using categorical post-stratification
4	GAM	f.year + s(doy) + dens + s(depth) + s(lon,lat)	26816.14	27331.48	Best fit
5	GAM	f.year + s(doy) + dens + s(depth)	26904.25	27373.29	without smoother over space
6	Zero-inflated GLMM	Positive catch: f.year + sc.doy + depth + lat + lon Zero catch: dens + lat + lon	27571.03	27754.14	Zero-inflated model did not fit better

Table 9. A list of tested spatial models for cod bycatch analysis and model selection (depth and doy, day of year, were incorporated with smoothing functions and dens, cod density, was included as a linear predictor). Model development followed incremental changes for optimization. AR1-first order autoregressive; RW-random walk; IID-independent and identically distributed; AIC-Akaike information criteria; BIC-Bayesian information criteria.

Model #	Spatial Effect	Spatiotemporal Effect	Covariates	AIC	BIC
1	on	AR1	depth, dens, doy	26366.45	26559.19
2	on	RW	depth, dens, doy	26403.25	26586.35
3	on	IID	depth, dens, doy	26370.13	26553.23
4	on	off	depth, dens, doy	26640.46	26813.93
5	off	IID	depth, dens, doy	26414.35	26587.82
6	on	IID	depth, doy	26374.34	26547.8
7	on	IID	doy, dens	26382.33	26546.16
8	on	IID	depth, dens	26447.78	26611.61
9	on	IID	doy	26432.71	26586.9
10	on	IID	dens	26462.91	26607.46
11	on	IID	none	26522.03	26656.94

Table 10. Proportion of missing data in commercial lobster fishing effort data for LFA 41 and the final predictions were scaled up using these proportions to impute missing information.

Year	Percentage of valid data
2016	0.96420582
2017	0.99940688
2018	1
2019	1
2020	0.99917253
2021	0.99874451
2022	0.99787836
2023	0.99933643
2024	0.99875854

Table 11. Cross-validated residual root mean square error (aggregated to the bathymetry grid to better represent model performance).ME-mean estimator; RF-Random Forest; GAM-generalized additive model; STM-spatio-temporal model.

Year	ME	RF	GAM	STM
2016	106.8 (0.16)	95.1 (0.79)	91.7 (0.69)	78.5 (1.35)
2017	100.4 (0.08)	91.8 (0.30)	91.3 (0.50)	63.6 (3.46)
2018	263.8 (0.08)	253.6 (1.08)	249.1 (2.27)	229.3 (11.05)
2019	55.2 (0.01)	54.1 (0.03)	54.8 (0.15)	52.0 (0.29)
2020	97.4 (0.02)	93.4 (0.16)	93.7 (0.21)	89.6 (0.53)
2021	290.8 (0.02)	290.8 (0.12)	288.3 (2.11)	288.1 (1.50)
2022	78.0 (0.02)	76.4 (0.09)	74.2 (0.11)	69.1 (0.43)
2023	87.5 (0.01)	84.1 (0.07)	83.7 (0.04)	77.2 (0.35)
2024	95.1 (0.01)	87.3 (0.19)	83.3 (0.50)	82.5 (0.27)
2016-2024	154.0 (0.01)	150.8 (0.15)	148.9 (0.72)	144.5 (1.48)
2018-2024	155.1 (0.01)	152.1 (0.16)	150.2 (0.73)	146.0 (1.51)

Table 12. Cross-validated residual mean absolute error (aggregated to the bathymetry grid to better represent model performance). ME-mean estimator; RF-Random Forest; GAM-generalized additive model; STM-spatio-temporal model.

Year	ME	RF	GAM	STM
2016	107.1 (0.17)	102.9 (0.33)	95.7 (0.48)	100.9 (0.60)
2017	46.3 (0.05)	42.9 (0.13)	48.9 (0.21)	39.7 (0.40)
2018	82.1 (0.03)	77.3 (0.16)	73.2 (0.22)	71.4 (0.61)
2019	26.0 (0.00)	26.3 (0.02)	33.3 (0.09)	24.8 (0.07)
2020	49.5 (0.01)	48.1 (0.06)	50.7 (0.11)	46.6 (0.15)
2021	32.0 (0.01)	32.1 (0.04)	35.3 (0.08)	30.3 (0.09)
2022	50.6 (0.01)	50.1 (0.04)	50.3 (0.07)	48.5 (0.12)
2023	59.6 (0.00)	57.7 (0.04)	55.7 (0.05)	55.8 (0.10)
2024	66.3 (0.00)	64.3 (0.04)	62.0 (0.09)	62.8 (0.10)
2016-2024	52.9 (0.00)	51.5 (0.02)	51.9 (0.03)	49.6 (0.05)
2018-2024	52.0 (0.00)	50.8 (0.02)	51.1 (0.03)	48.9 (0.05)

*Table 13. Predicted annual total bycatch (kg) using spatio-temporal model (STM) for 4X5Y (coefficient of variation is the ratio between standard deviations and estimates). *Predictions for 2016 and 2017 were not recommended for management consideration due to lack of data and high prediction uncertainty; prediction for 2024 did not include all efforts or bycatch within the year due to partial data. ME-mean estimator.*

Year	Total cod bycatch (kg) (STM)	Standard deviation	Lower Confidence Interval	Upper Confidence Interval	Coefficient of variation	Total cod bycatch (ME)	Total fishing effort (trap hauls)
2016*	1,880,786	20,817,708.8	25,482	138,815,441	11.1	1,795,357	32,584,884
2017*	287,261	1,210,097.21	10,023	8,233,080	4.2	702,141	29,891,246
2018	1,065,864	1,962,584.08	98,255	11,562,470	1.8	1,404,383	33,336,893
2019	461,325	205,018.852	200,737	1,060,198	0.4	424,297	32,257,729
2020	527,266	1,011,976.84	46,163	6,022,296	1.9	720,999	28,482,123
2021	576,267	744,424.729	82,655	4,017,719	1.3	511,701	31,525,358
2022	710,566	508,186.096	201,607	2,504,397	0.7	781,828	30,271,136
2023	830,836	283,160.983	433,851	1,591,073	0.3	912,776	29,841,533
2024*	606,980	838,248.021	80,135	4,597,520	1.4	617,940	18,139,940

*Table 14. Predicted annual total bycatch (kg) using spatio-temporal model for each lobster fishing area (LFA) within 4X5Y. The predictions were aggregated to the LFAs using model output that relied on the bathymetry grid; small discrepancies may arise when assigning a bathymetry grid to an LFA. *Predictions for 2016 and 2017 were not recommended for management consideration due to high uncertainty; prediction for 2024 did not include all efforts or bycatch within the year due to partial data.*

Year	Predicted annual total bycatch (kg) LFA 33	Predicted annual total bycatch (kg) LFA 34	Predicted annual total bycatch (kg) LFA 35	Predicted annual total bycatch (kg) LFA 36	Predicted annual total bycatch (kg) LFA 38	Predicted annual total bycatch (kg) LFA 41
2016*	775,512	995,883	14,230	38,657	47,567	8,617
2017*	152,692	116,479	1,596	4,923	6,321	5,246
2018	261,838	767,553	11,146	13,012	8,634	3,680
2019	120,480	313,507	5,474	10,433	10,727	703
2020	257,571	237,522	3,037	11,423	14,377	3,334
2021	259,874	273,109	7,130	20,787	14,030	1,335
2022	327,926	370,804	929	3,563	6,677	666
2023	349,375	466,317	3,012	6,374	5,297	461
2024*	193,378	355,415	1,659	28,564	25,429	2,530