

Bluefin Combined Index Model V1

Katie Lankowicz

07 July, 2023

Contents

1	Motivation	1
2	BFT data	2
2.1	Model domains	2
2.2	BFT size categories	3
2.3	Environmental covariates	4
2.4	US BFT Recreational Catch	4
2.5	Canadian BFT Commercial Landings	6
2.6	Data needs	8
3	VAST	9
3.1	Response variable and effort adjustment	9
3.2	Density and Catchability Covariates	10
3.3	Error structure	10
3.4	Spatial, temporal, and spatiotemporal effects	10
4	Preliminary results	11
4.1	Joint indices of abundance	11
4.2	Maps of spatial density	12
4.3	Spatial metrics and range shifts	13
5	Model caveats	14
6	GitHub	14
	References	14

1 Motivation

In the last several decades, global climate change has resulted in warming ocean temperatures and altered marine habitat suitability. Responses to these environmental changes can be seen at the individual, population, and ecosystem level for many marine fish species. Frequently, marine fish will shift their spatial-temporal distribution to track environmental conditions that are better-suited to their energetic requirements. It is suspected that the western stock

of Atlantic Bluefin Tuna (BFT) is currently undergoing a climate-forced range shift due to rapidly warming conditions in the Northwest Atlantic. Recent stock assessments for BFT in US and Canadian waters have provided conflicting trends in catch-per-unit-effort (CPUE) indices for BFT, where the Canadian fishery has seen an increase in BFT CPUE and the American fishery has seen a decrease.

This project will create a joint CPUE index for BFT across American and Canadian waters in the Northwest Atlantic using Vector Autoregressive Spatio-Temporal (VAST) modeling. VAST is a flexible model framework based on delta-generalized linear mixed models, and is capable of incorporating information from multiple sources and gear types to create a joint index. VAST can also incorporate the effects of habitat and catchability covariates on organism spatio-temporal distribution. VAST outputs include spatial metrics including spatio-temporal density maps, hierarchical clustering of spatial areas based on organism density, stock center of gravity, area of utilization, and quantile of range edges.

The remainder of this document will be used to report progress on cleaning and merging American and Canadian BFT catch data and making VAST model structure decisions.

2 BFT data

BFT data from American and Canadian fisheries are in variable formats and need to be heavily manipulated prior to merging. The goal is to create a dataframe where each row represents a sampling event. The following variables are desired for each sampling event:

- Unique trip identifier
- Date
- Spatial location (latitude and longitude, decimal degrees)
- Number of hours spent fishing
- Number of large BFT caught
- Sea surface temperature
- Water depth
- Sea level pressure
- Basin-wide daily NAO index
- Basin-wide monthly AMO index
- Wind direction and velocity
- Local chlorophyll-A concentration (as a proxy for plankton distribution)
- Local prey biomass

2.1 Model domains

Currently, the model's temporal domain is 1993 - 2020. Though BFT total catch data were recorded prior to 1993, information to determine the size of each fish was not reliably recorded. The final model will only include data for large BFT (see following section), and so data without associated length or weight information is not useful. The temporal domain currently ends at 2020 because that is the most recent year with Canadian catch data. American catch data ends at 2021. The temporal domain of the model could be extended to 2023 with updated information from both the American and Canadian fisheries.

The model spatial domain is the combined area of American and Canadian exclusive economic zones (EEZs) in the Northwest Atlantic, bounded on the south by Cape Hatteras, North Carolina and on the north by Cape North, Nova

Scotia. The model does not consider the Gulf of Saint Lawrence, nor is any tuna catch data from that area included. The model will estimate indices of abundance within both of the spatial strata– US EEZ waters and Canadian EEZ waters.

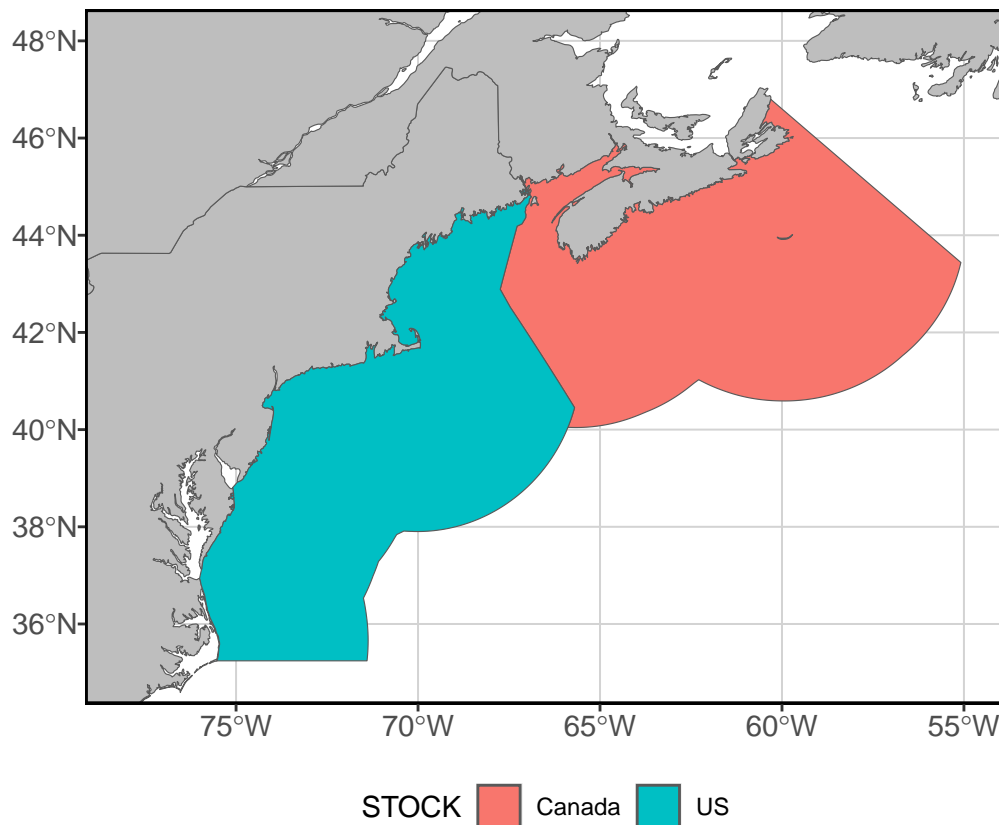


Fig. 1: VAST model spatial domain

2.2 BFT size categories

BFT are categorized into six size classes by both weight and length. BFT migrations and habitat use are expected to vary with age, and so the ideal model will generate separate indices and spatial metrics for different age groupings. There is not enough information to do this for each of the six size classes, and so the size class information was pooled into two groupings: small and large. Small BFT are 144 cm (56.7 in) and shorter, while large BFT are longer than 177 cm (69.7 in). This structure was chosen to match the structure of current US handline BFT indices. Because the “Small medium” size class straddles the line between groups, these fish have been excluded from analysis.

After initial data exploration, it was determined that too few small BFT were caught in Canada to generate robust indices of abundance. Therefore, this model will only consider large BFT– tuna which fall within the “large-medium” or “giant” size classes.

BFT were categorized according to length, preferentially. If no length information was recorded, weight was used to categorize. If neither length nor weight information was recorded for a fish, it could not be included in analysis.

Table 1: Bluefin tuna size classes

Size class	Grouping	Curved Fork Length	Pectoral Curved Fork Length	Weight (lbs)
Young school	Small	<27 in	<20 in	<14 lbs
School	Small	27 - <47 in	20 - <35 in	14 - <66 lbs
Large school	Small	47 - <59 in	35 - <44 in	66 - <135 lbs
Small medium	NA	59 - <73 in	44 - <54 in	135 - <235 lbs
Large medium	Large	73 - <81 in	54 - <60 in	235 - <310 lbs
Giant	Large	81+ in	60+ in	310+ lbs

2.3 Environmental covariates

Several ocean climate variables and static properties are expected to cause shifts in BFT distribution. For this model, we will include sea surface temperature, sea level pressure, basin-wide daily NAO, basin-wide monthly AMO, and water depth. When building the VAST model, these will be called density covariates.

2.4 US BFT Recreational Catch

Catch and effort information for the US recreational BFT fishery is recorded by the Large Pelagics Survey, which is a dockside/telephone survey of a random sample of private and charter boat captains targeting large pelagics. We subset the full LPS dataset so only data from trips targeting BFT were included. Other filters were also applied: only tuna landed in Virginia and northwards were included, fishing trips needed to be between 1 and 24 hours in length, and only fishing trips ending on June 1st through October 31st of each year were included.

LPS data recording procedures changed in 2002, and so data are stored across two sets of files (1993-2002, 2002-present). These needed to be cleaned and merged. This process can be viewed in the accompanying RMarkdown document. An example of results will also be provided.

Note that sea surface temperature and water depth are not always provided by the anglers. We have merged these spatiotemporal catch data with NOAA 1/4° Daily Optimum Interpolation Sea Surface Temperature (OISST) and GEBCO 15 arc-second gridded bathymetric data to get complete data coverage for SST and depth.

Table 2: American Large Pelagic Survey example

	Trip ID	Year	Month	Day	Hours fished	Lon	Lat	nCatch	SST (C)	Depth (m)	Pressure	NAO	AMO
1	1	1993	6	12	10	-74.91667	36.58333	0	20.64	-2.1	101955	0.1363429	-0.197
2	2	1993	6	12	10	-75.41667	36.75000	0	20.99	-2.1	101993	0.1363429	-0.197
5	3	1993	6	19	5	-75.08333	36.91667	0	22.29	-2.1	102005	2.6760383	-0.197
7	4	1993	6	20	7	-75.08333	36.58333	0	22.43	-2.1	101970	2.2424715	-0.197
8	5	1993	6	20	7	-75.25000	36.75000	0	22.60	-2.4	101970	2.2424715	-0.197
11	6	1993	6	20	9	-75.75000	36.91667	0	22.43	-1.8	101970	2.2424715	-0.197

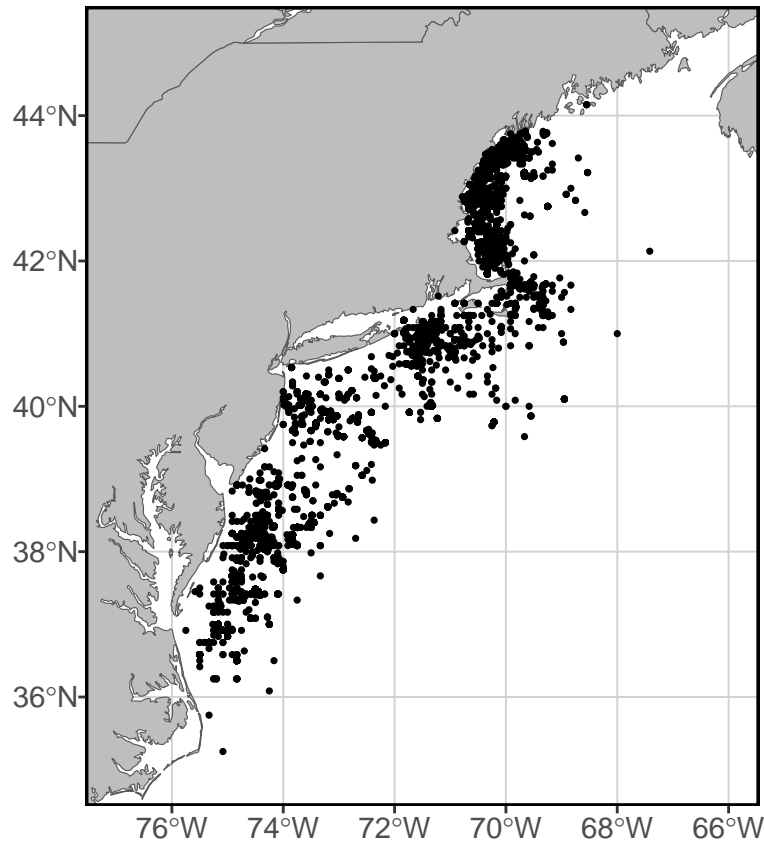


Fig. 2: Spatial distribution of American recreational catch

2.5 Canadian BFT Commercial Landings

Catch and effort data for the Canadian commercial BFT fishery are recorded in commercial logbooks. We will only be working with commercial data from the Southwest Nova Scotia region, though data exists for the Gulf of Saint Lawrence region as well. The same data filters regarding fishing hours and month fished applied to the US data were applied to the Canadian data. Though the US recreational fishery is likely all rod and reel, the Canadian commercial fishery also includes harpoon and tended line gear. This could later inform some model structure decisions.

Data collection methods changed in 2003, and so the data are stored across two sets of files (Commercial Landings 95-03, Historical Data). These needed to be cleaned and merged. This process can be viewed in the accompanying RMarkdown document. An example of results will also be provided.

There are some concerns with the historic (1993-2002) Canadian data. The data have several columns that I cannot make sense of without metadata. Although I assume I have only been provided with BFT catch, there is a column for species landed in the data. The values within this column are numeric (252-256), and I do not know what these numbers mean. For now, I have made the assumption that all these species codes refer to BFT. In the future, I can subset out non-BFT data if I am made aware that some of these species codes are not BFT.

Table 3: Canadian commercial data example

Trip ID	Year	Month	Day	Hours fished	nCatch	SST (C)	Depth (m)	Pressure	NAO	AMO	Lon	Lat
1	1993	8	1	10	0	15.17	-673.3	101355	1.5409321	-0.245	-65.55000	42.05000
2	1993	8	1	12	0	15.77	-1088.9	101355	1.5409321	-0.245	-65.46667	42.03333
3	1993	8	1	10	0	15.17	-378.1	101355	1.5409321	-0.245	-65.60000	42.05000
4	1993	8	1	10	0	15.17	-378.1	101355	1.5409321	-0.245	-65.60000	42.05000
5	1993	8	6	10	0	15.62	-558.3	101590	0.3863121	-0.245	-65.56667	42.08333
6	1993	8	7	10	0	15.83	-649.6	101773	0.2472577	-0.245	-65.53333	42.03333

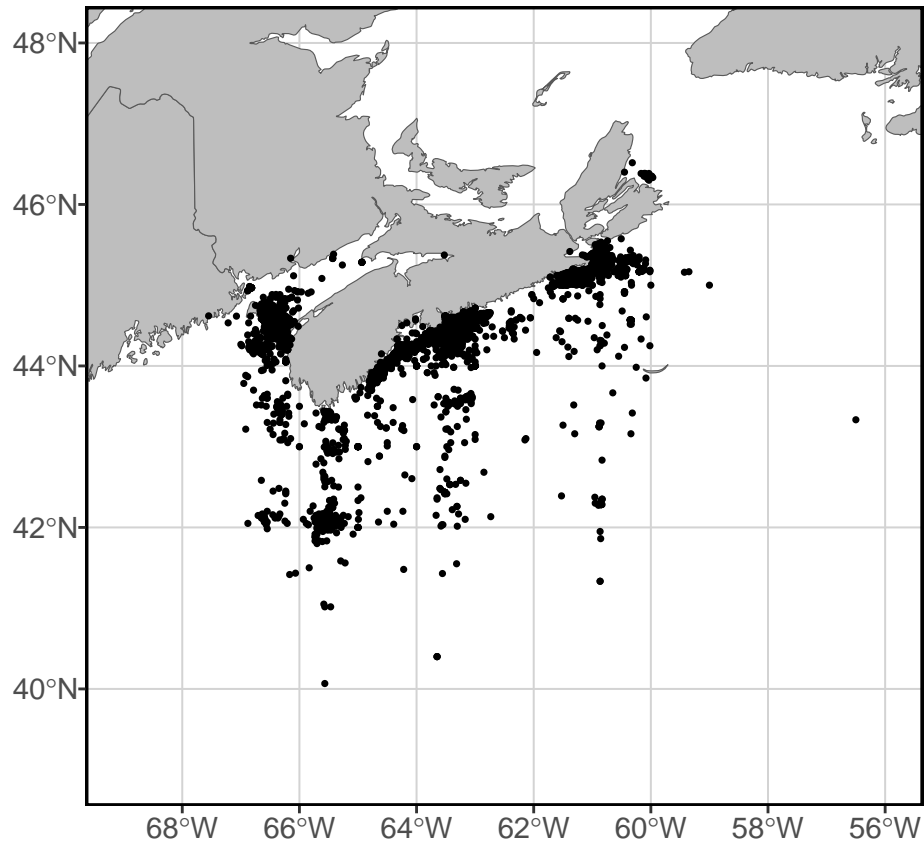


Fig. 3: Spatial distribution of Canadian commercial catch

2.6 Data needs

There are a few environmental variables that have not yet been incorporated into the resulting data. I have created a VAST model of prey distribution based on NOAA herring, menhaden, and mackerel observer data. However, these data do not have 0 values (an important part of a VAST model) and do not have any coverage in Canadian waters. I can work with catch data that only have positive catch (no 0s), but it would be best to fill out the data with information from other sources to get better models of prey spatiotemporal density.

Chlorophyll was noted as a potential covariate for BFT distribution, but I have not yet found a data source that has complete coverage in our spatial and temporal domains. VAST models cannot tolerate missing values for density covariates, and the many gaps in MODIS data caused by cloud coverage are currently preventing me from including chlorophyll data. Suggestions are welcomed.

Wind direction and velocity have not yet been incorporated. Ideally, we would have information on the wind direction and velocity at the exact space and time of each tuna observation, though I could extract this information from data with larger spatial grain. I have spent less time looking for a source for these data. Suggestions are welcomed. I can be reached at klankowicz@gmri.org

Table 4: Data needs

Data source	Spatial area	Years
Prey data (mackerel, herring, menhaden)	Canadian waters in spatial domain	All model years
Prey data (mackerel, herring, menhaden) with non-encounters	American waters in spatial domain	All model years
Chlorophyll distribution (must have 100% coverage)	Entire model spatial domain	All model years
Wind direction	Entire model spatial domain	All model years
Wind velocity	Entire model spatial domain	All model years
American recreational catch data	American waters in spatial domain	2022-2023
Canadian commercial catch data	Canadian waters in spatial domain	2021-2023

3 VAST

VAST models were used to estimate BFT spatial density over time and create joint indices of abundance using all available American and Canadian catch data. VAST is a framework for implementing spatial delta-generalized linear mixed models (delta-GLMM) and can be manipulated to provide estimates for multiple categories of interest and spatial strata Thorson (2019). It is structured to utilize two linear predictors; the first linear predictor estimates encounter probability, and the second linear predictor estimates catch rates. The first linear predictor can be represented as

$$\rho_1(i) = \beta_1(c_i, t_i) + \omega_1^*(s_i, c_i) + \varepsilon_1^*(s_i, c_i, t_i) + \eta_1(v_i, c_i) + \nu_1(c_i, t_i)$$

where $\rho_1(i)$ is the predictor for observation i for category c_i at location s_i and time t_i . $\beta_1(c_i, t_i)$ represents temporal variation for each category and time, $\omega_1(s_i, c_i)$ represents spatial variation for each location and category, $\varepsilon_1(s_i, c_i, t_i)$ represents spatiotemporal variation for each location, category, and time, $\eta_1(v_i, c_i)$ represents vessel effects for each vessel and category, and $\nu_1(c_i, t_i)$ represents the effect of density covariates for each category and time. The second linear predictor is structured the same way. Both linear predictors incorporate fixed and random effects, and spatial and spatiotemporal variation are approximated as Gaussian Markov random fields Thorson (2019).

Implementation of VAST models requires several structural and data inclusion decisions, as outlined in Thorson (2019). Decisions used in this effort will be discussed in the following subsections.

3.1 Response variable and effort adjustment

Our data provide information on the number of fish of each size group caught during each fishing trip, as well as the number of hours fished on that trip. Gear type varies and this likely affects catchability and actual effort. For now, catch will be adjusted for effort by dividing the catch by number of hours fished.

VAST models require an input for effort to scale the response variable. For most fishery-independent surveys, this input would be similar to area swept. For this fishery-dependent dataset utilizing multiple gears, we have already accounted for effort in our manipulation of the response variable. Therefore, we will assign effort to be a unitless 1 for all observations.

3.2 Density and Catchability Covariates

VAST allows for the effects of both density and catchability covariates to be included in modeling efforts. Catchability covariates are processes one would expect to affect the ability to observe the target organism without necessarily affecting the distribution of the organism. Density covariates are processes that directly affect the distribution of the target organism, regardless of ability to observe it. Both covariates affect the catch rate of the target organism, but only density covariates are used to predict target organism density within the spatial domain. Therefore, VAST “controls for” catchability covariates and “conditions on” density covariates. VAST is unable to distinguish whether potential covariates should be treated as catchability or density covariates; this must be decided with theoretical insight from an analyst.

In this model, we will utilize the environmental covariates and basin-scale climate indices as density covariates. Density covariates were chosen after a model selection process to determine their utility to describe patterns in BFT spatiotemporal density. Currently, a simple linear model is used to describe the effects of these covariates on BFT CPUE. In the future, other models will be tested; it is likely that the best model will be more fluid, like a generalized additive model using polynomial splines.

We will not explicitly include catchability covariates, but will allow for variation in catchability between the American and Canadian fisheries by including them as “vessel effects.”

3.3 Error structure

As recommended by model developers for adjusted-effort models, a delta-lognormal distribution was chosen for the first linear predictor and an alternative “Poisson-link delta-model” was chosen for the second linear predictor.

3.4 Spatial, temporal, and spatiotemporal effects

Spatial, temporal, and spatiotemporal autocorrelation can be included in both linear predictors. A model selection process was used to justify the use of spatial and spatiotemporal random effects in the first and second linear predictors. The intercept for each linear predictor was defined as a fixed effect for each time step— this ensures independent estimates of abundance for each time step, which is most appropriate for creating abundance indices (Thorson 2019). Instead, a temporal component was estimated for the spatiotemporal components in both linear predictors. This is recommended for indices generated by multiple data sources that do not necessarily sample the same locations in every time step (Thorson 2019). Without this estimation, unrealistic “hot spots” may develop or be carried through the time series when this is inappropriate. The use of a temporal component for spatiotemporal components can also help interpolate density in unsampled time steps (Thorson 2019). In this model, temporal correlation of these spatiotemporal variations was set as an AR1 process.

4 Preliminary results

4.1 Joint indices of abundance

Indices of abundance for the early period in the model's temporal domain (1993-2000) are highly variable and look to have some issues due to data quality. More work will need to be done to better account for the lack of observations with 0 tuna in this time period.

Model outputs indicate that large BFT CPUE is high and increasing in Canadian waters, and generally has been since the early 2000s. Large BFT CPUE is much lower in American waters, and has been generally stable since the early 2000s. The large BFT population is driven by fluctuations in the Canadian portion of the stock.

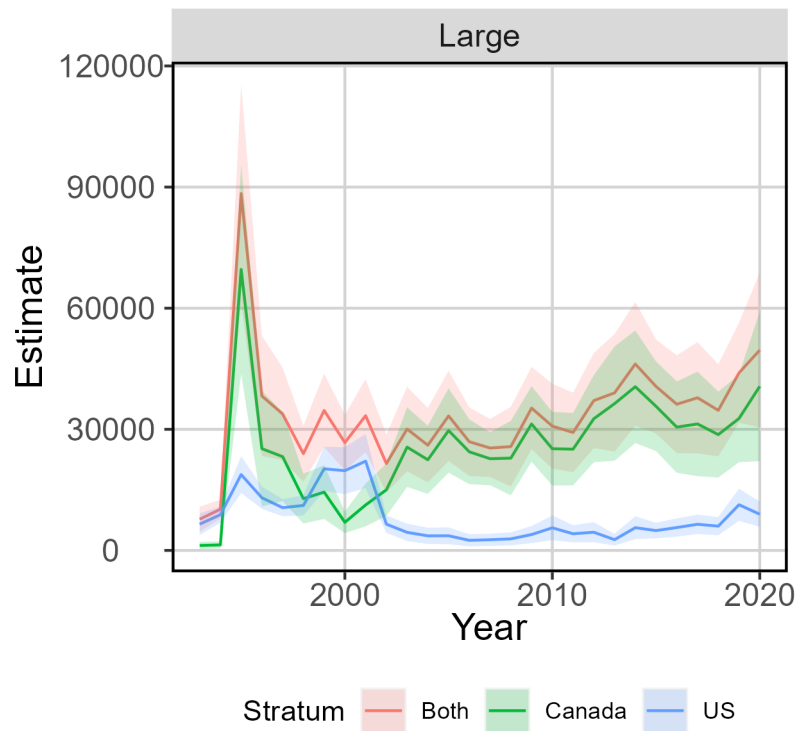


Fig. 4: Joint indices of abundance

4.2 Maps of spatial density

Typically, large BFT density is highest in Canadian waters. The early period of the model indicated areas of high density near Cape Cod, though this is likely caused by the lack of observations with 0 captured tuna prior to 2003. After 2003, large BFT density was highest on eastern Georges Bank and the Scotian Shelf. Density of large BFT south of Cape Cod has declined dramatically throughout the modeled period.

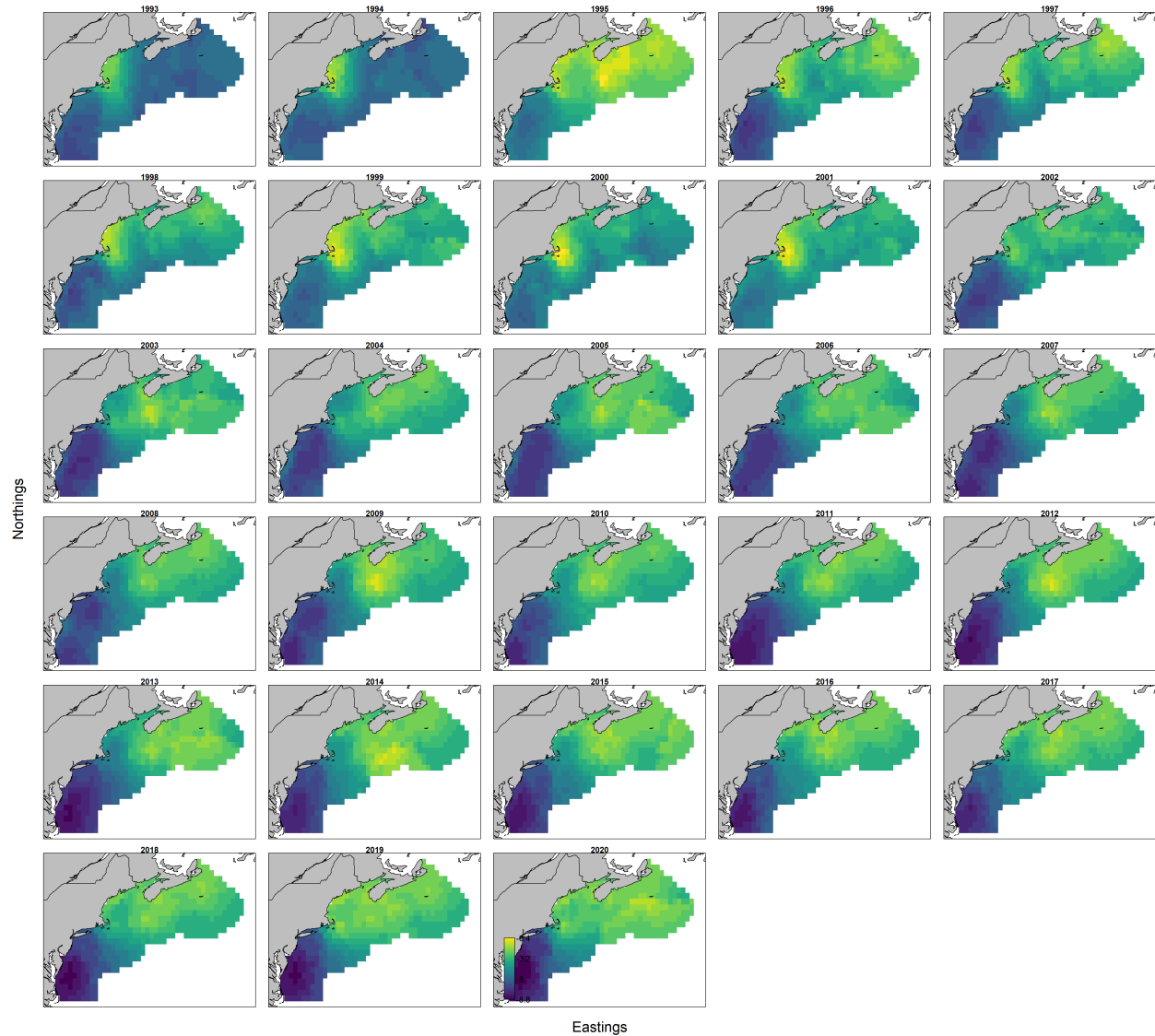


Fig. 5: Large BFT spatio-temporal density

4.3 Spatial metrics and range shifts

The model needs to be run once for each spatial strata in order to generate spatial metrics for each strata. As of right now, it takes approximately 9 hours to run on a laptop. Spatial metrics have only been generated for the Canadian stock.

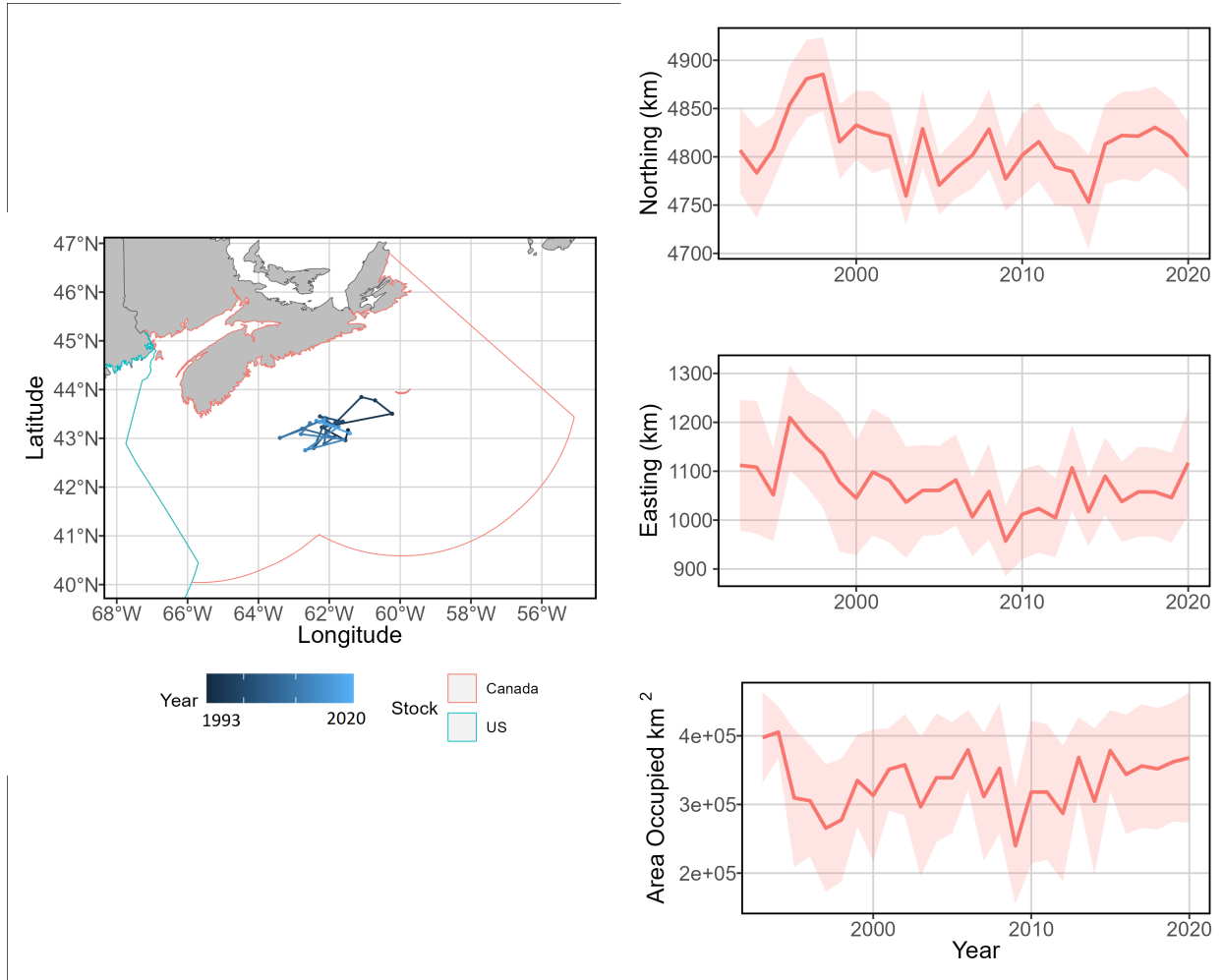


Fig. 6: Spatial metrics for large BFT in Canadian EEZ

The center of gravity of the large BFT stock has remained fairly consistent over the modeled period. Earlier years saw a slightly more north-and-east distribution, but since the early 2000s large BFT distribution has remained relatively constant. The spatial area used by the population has seen a slight increase since the early 2000s, indicating a decrease in average spatial density.

5 Model caveats

Both the American and Canadian datasets report very few sampling events with 0 catch. In the American dataset, the 1993-2002 period has no zeros reported. This likely skews the American indices to incorrectly report high BFT abundance in this period, despite the AR1 temporal correlation process.

There is clear age-based spatial structure of BFT (smaller BFT prefer warmer southerly waters, larger BFT prefer colder northerly waters). Very few small BFT were caught in Canadian waters, and therefore there is not enough information for the model to predict and interpolate BFT abundance in this area. This is not a barrier that can be overcome with more data. Abundance indices between US and Canadian waters can only be reliably created for the large size class.

6 GitHub

To follow model progress live, please [visit the Github repository](#) for the project.

References

- Thorson, J. T. 2019. [Guidance for decisions using the Vector Autoregressive Spatio-Temporal \(VAST\) package in stock, ecosystem, habitat and climate assessments](#). Fisheries Research 210:143–161.
- Thorson, J. T., and L. A. K. Barnett. 2017. [Comparing estimates of abundance trends and distribution shifts using single- and multispecies models of fishes and biogenic habitat](#). ICES Journal of Marine Science 74(5):1311–1321.
- Thorson, J. T., R. Fonner, M. A. Haltuch, K. Ono, and H. Winker. 2017. [Accounting for spatiotemporal variation and fisher targeting when estimating abundance from multispecies fishery data](#). Canadian Journal of Fisheries and Aquatic Sciences 74(11):1794–1807.
- Thorson, J. T., H. J. Skaug, K. Kristensen, A. O. Shelton, E. J. Ward, J. H. Harms, and J. A. Benante. 2015. [The importance of spatial models for estimating the strength of density dependence](#). Ecology 96(5):1202–1212.