

Atlantic Cod VAST model Technical Document

Katie Lankowicz

DRAFT 19 April, 2023

Motivation

Atlantic cod (*Gadus morhua*) are an ecologically, economically, and culturally critical part of New England's groundfish fishery. However, the fishery has been under a disaster declaration since 2012 due to historically low abundance and rapidly declining stocks. The Northeast Fishery Science Center's (NEFSC) twice-annual bottom trawl survey has been an important tool to assess groundfish stocks from Cape Hatteras, NC to Nova Scotia since the early 1960s. Through its 60+ year history, the survey has mostly sampled offshore areas deeper than 18m and with relatively smooth and soft bottom habitat. Though the survey has wide spatial and temporal coverage, there is concern among fishers that it does not adequately sample areas of high habitat complexity (cobble or boulder fields) that likely host dense populations of cod, particularly younger age classes. Many state and academic institutions conduct annual trawl surveys that may help to fill data gaps for groundfish stocks in shallower inshore waters. In 2016, the NEFSC launched a cooperative bottom longline survey to explore groundfish abundance and distribution in areas with high-complexity bottom habitat.

The goal of this project is to build a joint index of abundance and map spatial density for Atlantic cod using all relevant state and federal groundfish survey data. This goal will be addressed using Vector Autoregressive Spatio-Temporal (VAST) models. VAST estimates spatial density of multiple categories of a target organism conditioned on "density covariates" and controlling for "catchability covariates." We will utilize bottom habitat substrate type and rugosity data, among other habitat and environmental covariates, to explore patterns in cod habitat use. We can further examine spatial overlap between groundfish survey operations and groundfish commercial fishing operations. Products from this modeling effort will be provided to the Atlantic Cod Stock Assessment Research Track working group.

Survey Data

Eleven surveys of groundfish abundance are currently available for use. Overall spatial coverage runs along the coast from Lubec, Maine to Cape Hatteras, North Carolina and spans the four defined cod spatial stock areas: Eastern Gulf of Maine, Western Gulf of Maine (spring and winter spawners), Georges Bank, and Southern New England. Overall temporal coverage spans from 1959 - 2022. Included surveys are:

- NEFSC Cooperative Bottom Longline Survey (longline)
- Eastern Gulf of Maine Sentinel Survey (jigging)
- SMAST Video Trawl Survey (combined trawl-camera)

- NEFSC Bottom Trawl Survey (trawl)
- DFO Bottom Trawl Survey (trawl)
- University of Rhode Island GSO Trawl Survey (trawl)
- Massachusetts DMF Industry-based survey (trawl)
- Massachusetts DMF Inshore Trawl Survey (trawl)
- Maine-New Hampshire Inshore Trawl Survey (trawl)
- Rhode Island DEM Trawl Survey (trawl)
- ASMFC Shrimp Trawl Survey (trawl)

These surveys utilize different gears, have variable spatial extents, are taken over multiple temporal periods, and are of variable temporal length (as in, number of years in which the survey has occurred). All surveys report number of cod and total weight (kg) of cod caught per tow, and some surveys process all or a portion of the catch to provide further biological detail.

Spatial extent

The NEFSC bottom trawl survey extends through Cape Hatteras, North Carolina, but the spatial range of Atlantic cod is likely further north of this point. For the purposes of this model, the spatial extent of interest will be American waters on the continental shelf from the northern edge of the Gulf of Maine through the mouth of the Chesapeake Bay.

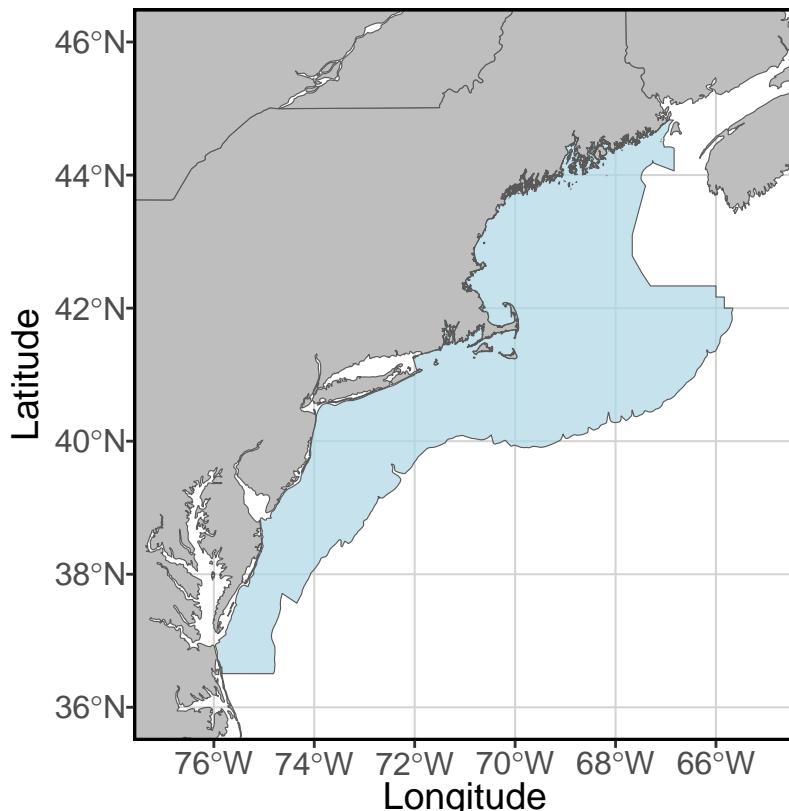


Figure 1: Spatial extent for cod density models

Strata

There is evidence that the relationships between various environmental covariates and cod spatial density vary among what are known to be distinct cod biological stock spatial areas (see Linner & Chen 2022). The model will therefore be run with each stock area as separate spatial strata—metrics of spatial dynamics, indices of abundance, and habitat associations can later be described for each stock area.

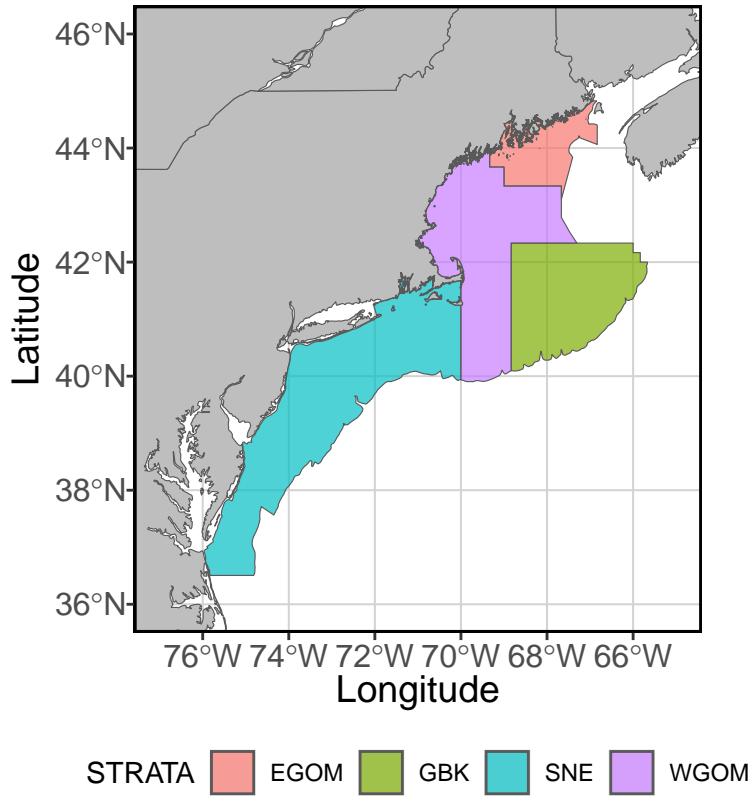


Figure 2: Cod stock areas

VAST input

VAST requires spatial gridpoints to which it will later extrapolate organism density. This process will be outlined below.

```
# Load shapefile of desired extrapolation region
shpsplit <- readOGR(here("Data/GIS/cod_all_strata.shp"))
head(shpsplit@data)

# Load again, strip strata
shp.all <- readOGR(here("Data/GIS/cod_region_wgs.shp"))
head(shp.all@data)
shp.all@data <- dplyr::select(shp.all@data,
                                OBJECTID, area_km)
colnames(shp.all@data) <- c('STRATA', 'area_km')
shp.all@data$STRATA <- 'ALL'
```

```

# Combine
shp <- rbind(shpsplit, shp.all)

# Transform to unprojected lat-lon
sps <- spTransform(shp,
                   CRS("+proj=longlat +datum=WGS84 +no_defs +ellps=WGS84"))

# Find UTM
lon <- sum(bbox(sps)[1,])/2
utmzone <- floor((lon + 180)/6)+1
crs_LL <- CRS('+proj=longlat +ellps=WGS84 +no_defs')
sps@proj4string <- crs_LL

# Convert the final in polygon to UTM
crs_UTM <- CRS(paste0("+proj=utm +zone=", utmzone, " +ellps=WGS84 +datum=WGS84
                        +units=m +no_defs"))
region_polygon <- spTransform(sps, crs_UTM)
region_polygon_sf <- st_as_sf(region_polygon)

# Construct the extrapolation grid for VAST using sf package
# Size of grid **in meters** (since working in UTM). Controls
# the resolution of the grid.
cell_size <- 1000

# This step is slow at high resolutions
region_grid <- st_make_grid(region_polygon_sf,
                            cellsize = cell_size,
                            what = "centers")

# Convert region_grid to Spatial Points to SpatialPointsDataFrame
region_grid <- as(region_grid, "Spatial")
region_grid_sp <- as(region_grid, "SpatialPointsDataFrame")

# combine shapefile data (region_polygon) with Spatial Points
# (region_grid_spatial) & place in SpatialPointsDataFrame data
# (this provides you with your strata identifier (here called
# Id) in your data frame)
region_grid_sp@data <- over(region_grid, region_polygon)
table(region_grid_sp@data$STRATA)
pulldat <- region_grid_sp
pulldat$STRATA[!is.na(pulldat$STRATA)] <- 'ALL'

```

```

test <- rbind(region_grid_sp, pulldat)
table(test@data$STRATA)
nrow(pulldat@data) + nrow(region_grid_sp@data) == nrow(test@data)

#region_grid_sp <- subset(region_grid_sp, is.na(region_grid_sp@data$OBJECTID)==FALSE)

# Convert back to lon/lat coordinates as that is what VAST uses
region_grid_LL <- as.data.frame(spTransform(test, crs_LL))
region_df <- with(region_grid_LL,
                    data.frame(Lon=coords.x1,
                               Lat=coords.x2,
                               STRATA,
                               Area_km2=( (cell_size/1000)^2),
                               row=1:nrow(region_grid_LL)))

# Filter out the grid that does not overlap (outside extent)
region <- subset(region_df, is.na(STRATA)==FALSE)

```

The outcome is a dataframe with spatial locations for the center of each grid cell, the strata in which the grid cell exists, and the spatial area of the grid cell in square kilometers.

```

## 'data.frame':    445342 obs. of  5 variables:
## $ Lon      : num  -74.9 -74.9 -74.9 -74.9 -74.9 ...
## $ Lat      : num  36.5 36.5 36.5 36.5 36.5 ...
## $ STRATA   : Factor w/ 5 levels "ALL","EGOM","GBK",...: 4 4 4 4 4 4 4 4 4 ...
## $ Area_km2: num  1 1 1 1 1 1 1 1 1 ...
## $ row      : int  93 94 95 96 97 98 99 100 976 977 ...

```

Temporal extent

Though some surveys were initiated in the 1950s and 60s, the time period of interest to this study begins in 1982. Some studies have provided full data from 2022, and others have not. For this reason, the model will cease after 2021.

Seasonality

Time steps in the model will not be annual, they will be seasonal. “Spring” months will be March through August, “Fall” months will be January, February, and September through December. This distinction is made to capture the seasonal spawning movements of adult cod.

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, catchability covariates will not be used. Instead, the effects of differences in sampling design on ability to observe cod will be included as vessel effects. This will be discussed in more detail in the Model Structure section.

Potential density covariates

The model will test the effects of several potential density covariates. As explained earlier, these are environmental and habitat characteristics that are expected to shape the distribution of cod.

Sediment data

Probability of mud, sand, gravel, cobble, and rock substrate existing in 1km by 1km grid cells through the entire continental shelf was provided by Brad Harris and Felipe Restrepo at Alaska Pacific University. These data are a expansion on SASI data, which previously provided polygons of distinct sediment grain types within the same spatial extent.

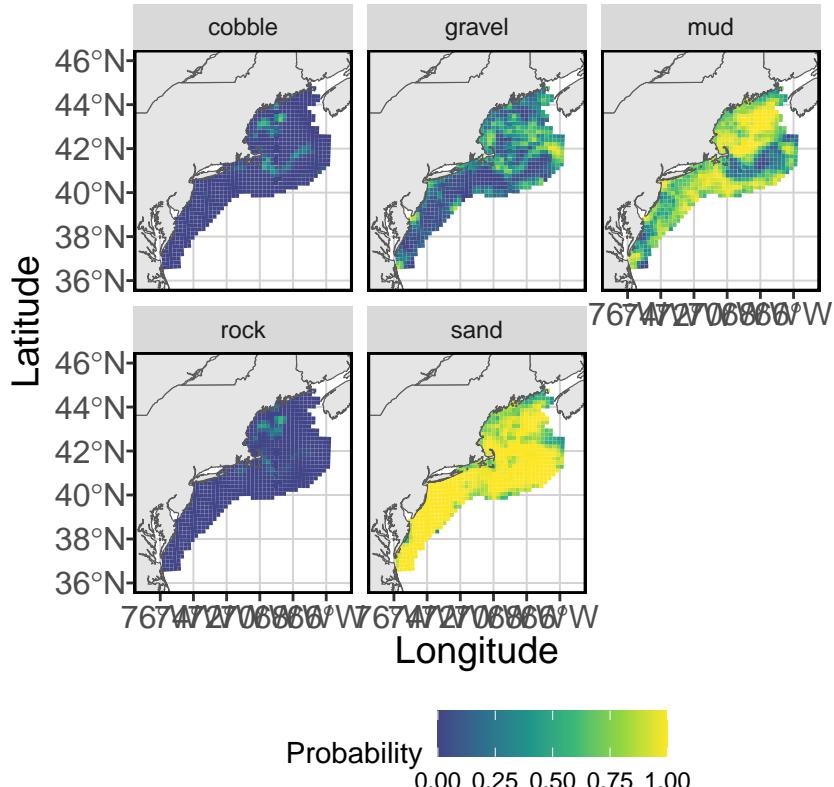


Figure 3: Sediment probability

Rugosity

Rugosity data comes from a dataset created by Kevin Friedland. Rugosity is calculated for grid cells with a 15 arc-second resolution. For visualization purposes, rugosity is then collected into two categorical variables – smooth and rough. Smooth habitat represents the area within each NEFSC bottom trawl survey strata that has a rugosity value below the 70th percentile for that strata. Rough habitat represents the 70-100th percentile rugosity area within the strata. In the model, the continuous and unitless value of rugosity is used as the input.

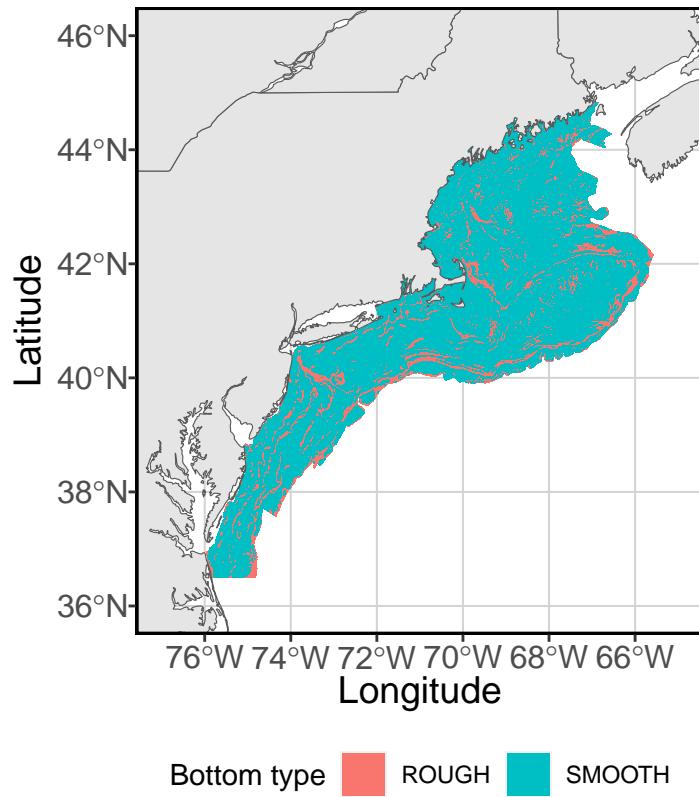


Figure 4: Rugosity

Bathymetry

Most surveys recorded depth at which the various gears were deployed. However, this model uses maximum water depth at the point of observation as a spatial covariate. Depth was interpolated from GEBCO 15 arc-second bathymetry.

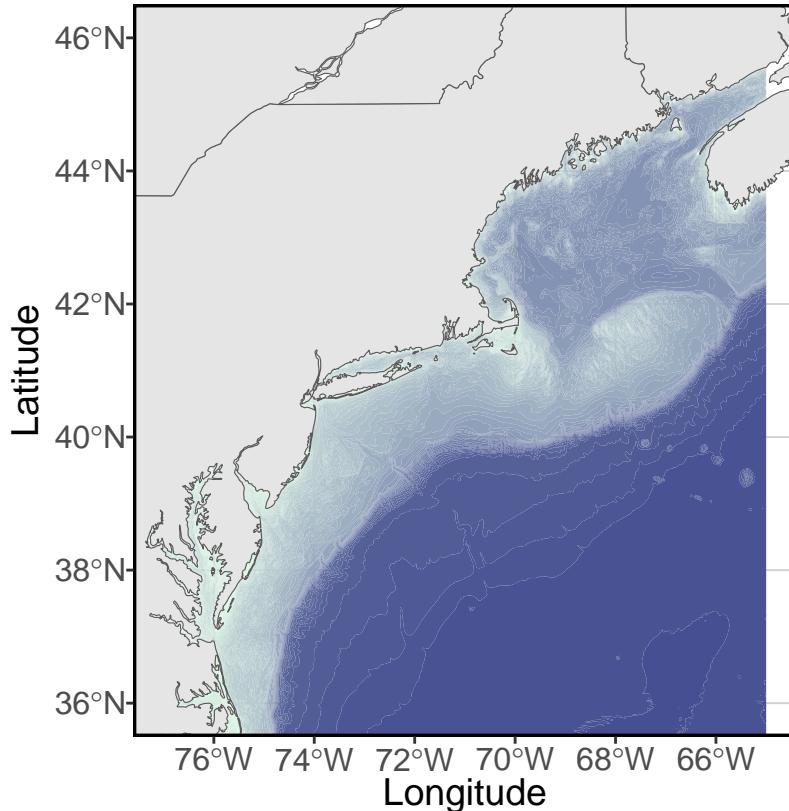


Figure 5: Bathymetry

Sea Surface Temperature

The previous density covariates were temporally stationary. The following measures of temperature are temporally dynamic. For sea surface temperature, OISST values were pulled from NOAA data sources and extracted at observation locations. OISST values were compared to field measurements, when available, and the two were found to be generally similar. The following plot describes overall differences between field-measured SST and OISST when both measures were available for the same observation.

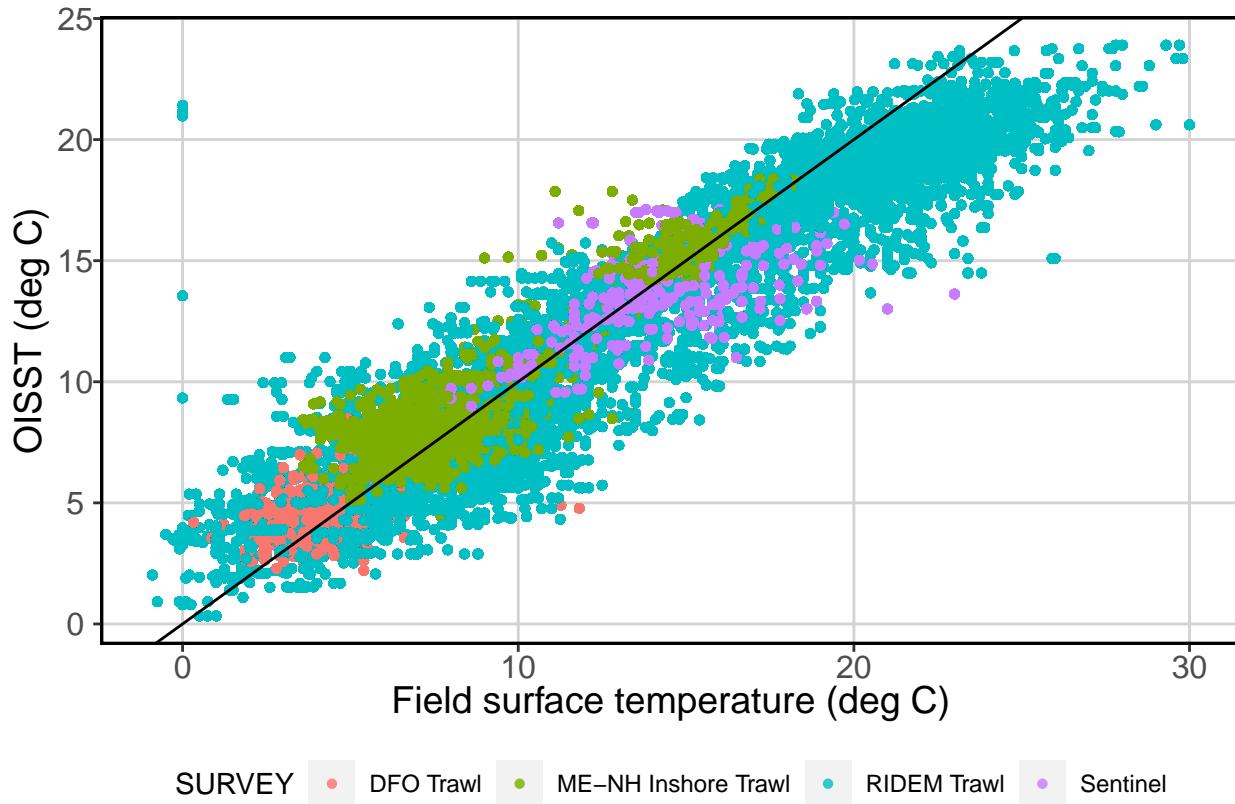


Figure 6: Comparison of field-measured SST to OISST

Bottom Temperature

Bottom temperature data were provided by Hubert du Pontavice. Bottom temperature within approximately 5' by 5' grid cells (in decimal degrees) were calculated at a daily timestep for 1982 to 2020. Because cod observation data were collected through 2021, bottom temperature data for this year were simply copied from 2020. This may not be accurate, and it may be better to exclude 2021 data in future efforts. Temperature data were provided as a conglomerate NCDF file for 1959-2020 and had some data wrapping errors- meaning, from 1993 onwards, results that should have made up the eastern edge of the spatial range were instead the western edge. This problem was fixed. Further, the bottom temperature product did not extend to the inshore strata included in the model. To fix this issue, bottom temperature was interpolated to the inshore strata from the existing data.

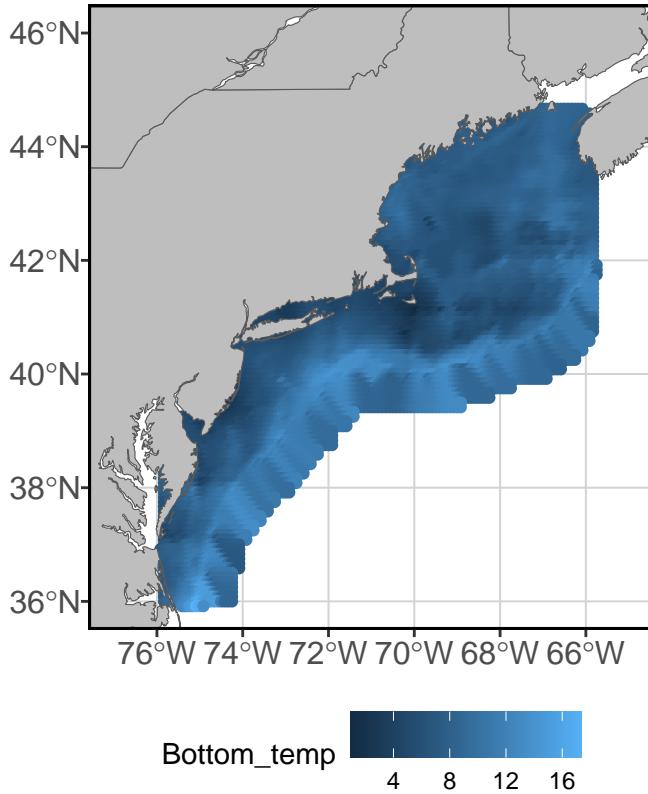


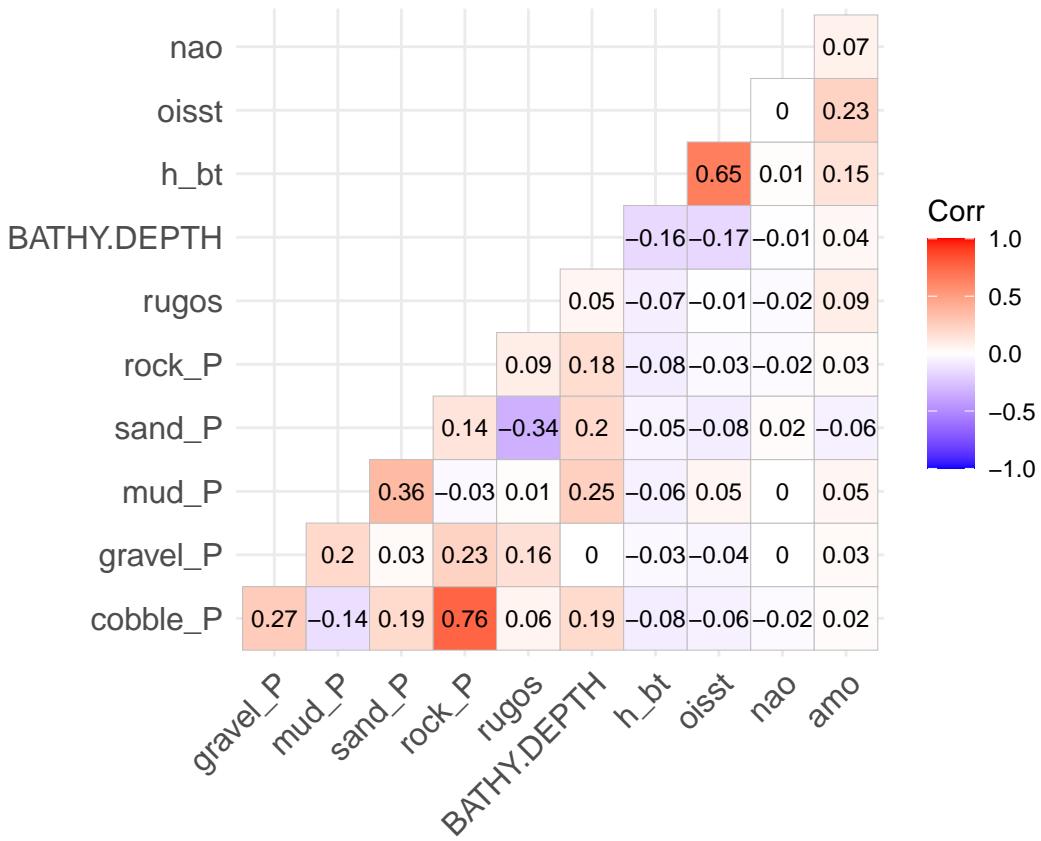
Figure 7: Bottom temperature on 1 January 1982

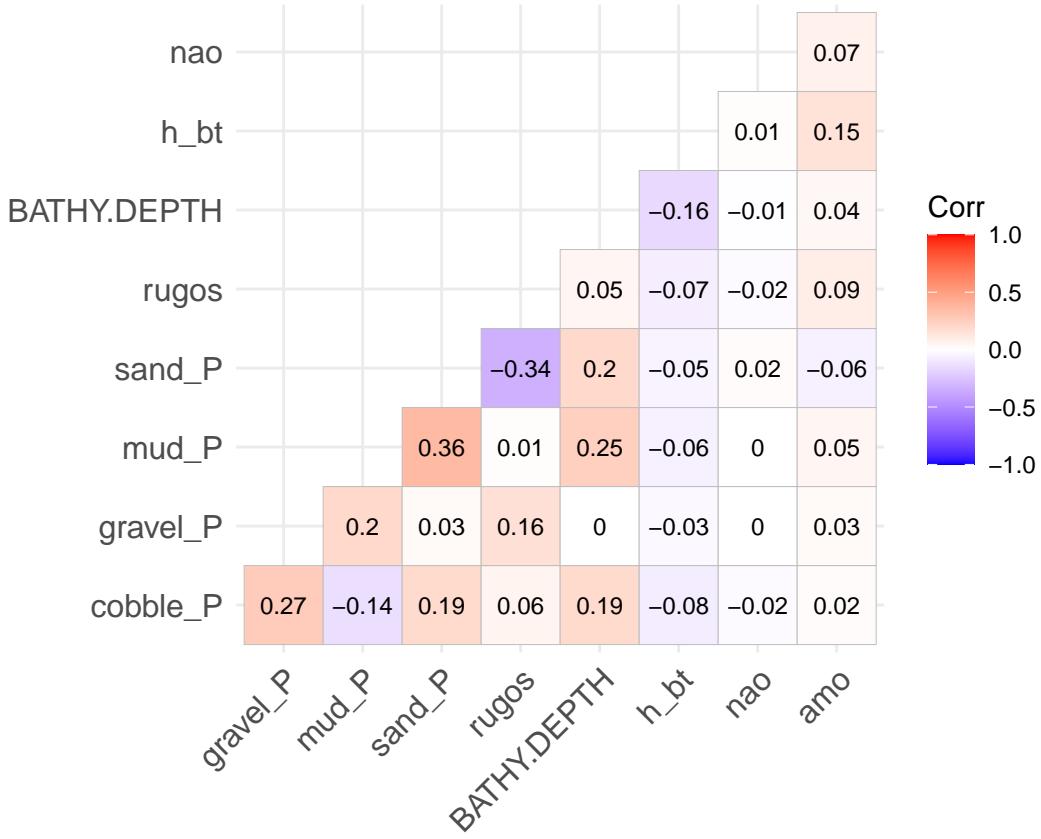
Climate Indices

Several climate indices have been demonstrated to have a relationship to cod abundance and distribution. Previous works have shown Gulf Stream Index (GSI) to be a particularly strong identifier. For this model, we have access to high temporal-scale NAO and AMO data. NAO data are provided at a daily timestep within the model, whereas AMO was provided at a monthly timestep. Both were basin-wide indices, meaning the value does not change between spatial locations within the region.

Collinearity

Potential spatial covariates were tested for collinearity. High correlations were found between two pairs: cobble and rock probability within 1km grid cells, and OISST and bottom temperature. Probability of rock and OISST were removed as covariates. For the first pair, rock was removed as bottom sample grabs (the bulk of sediment samples that make up the dataset) are more likely able to accurately characterize cobble distribution than larger boulders. For the second pair, OISST was removed as cod are groundfish and therefore more likely to be affected by bottom temperature than surface temperature. Once these two collinear relationships were addressed, the remainder of density covariates were not correlated and therefore were included in the model.





Covariate selection

Density covariates were selected for inclusion into the final model by running a subset of the original observation data (last 5 years of NEFSC bottom trawl data) with some combination of variables as spatial density inputs, then comparing models using AIC. The model with the lowest AIC was the model that included spatial density terms for all potential covariates.

```
modname deltaAIC fixedcoeff randomcoeff use_anisotropy omega1 omega2 epsilon1 epsilon2 beta1 beta2 AIC
```

Model settings

Model settings were also explored using model selection. Models were tested for different values in the Field Configuration (testing the use of spatial and temporal variation), anisotropy, and overdispersion. The best model used anisotropy, spatial and temporal effects.

Settings selection

```
modname deltaAIC fixedcoeff randomcoeff use_anisotropy omega1 omega2 epsilon1 epsilon2 beta1 beta2 AIC converged2
```

Distribution functions

Error distributions for the first and second linear predictors (presence and abundance, respectively), were set depending on the advice of the author. The author suggests using delta-lognormal distribution for the first linear predictor and alternative Poisson-link delta models for density within each patch for catch (abundance or count) data.

Knots and maximum cells

It has been said that selection of knots and grid cells is “an art, not a science.” For now, 200 knot locations will be used to make the SPDE grid on which VAST is run. Default number of grid cells on which data are extrapolated is 2000. In or VAST region, we have made more than 400,000 grid cell locations; for now, we will stick to the default of 2000 locations.

Final model

Model inputs

Age/Size categories

Vessel effects

Effort

Results