A species distribution model (SDM) approach to representing anchovies in the NWACS ecosystem model

Micah Dean and Mike Celestino

SEDAR102-RW-05

25 July 2025

Please cite this document as:

Dean, Micah and Mike Celestino. 2025. A species distribution model (SDM) approach to representing anchovies in the NWACS ecosystem model. SEDAR102-RW-05. SEDAR, North Charleston, SC. 14 pp.

SEDAR 102 WP-05: A species distribution model (SDM) approach to representing anchovies in the NWACS ecosystem model

A working paper presented to the 2025 ASMFC benchmark assessment of ecosystem reference points for Atlantic menhaden

Micah Dean – Massachusetts Division of Marine Fisheries
Mike Celestino – New Jersey Division of Fish and Wildlife

January 22[th], 2025

Introduction

Atlantic coast anchovies (*Anchoa mitchilli*, *Anchoa hepsetus, Engraulis eurystole*) are small pelagic forage fishes that are typically most abundant close to shore yet seasonally can be found up to 20 km from land. In addition, the population biomass and latitudinal extent of their distribution varies seasonally. As such, it is difficult to obtain an accurate index of abundance or an estimate of population biomass by calculating an empirical average of state and federal trawl surveys, each of which captures only a fraction of the population's distribution and at different times of year. This paper describes the use of a species distribution model (SDM) to represent Atlantic coast anchovy populations, accounting for their spatial-seasonal distribution patterns.

Methods

Survey Data

A total of 8 trawl surveys relevant to Atlantic coast anchovy populations were identified (Table 1). Collectively, these surveys provided 59,119 observations of anchovy catch per unit effort (CPUE – kg/tow) within the NWACS model domain for the years 1985-2023 (Figures 1 & 2). A common set of candidate variables with potential relevance to anchovy abundance was assembled across all datasets (Table 2). Examination of these raw survey data revealed that anchovies were most associated with depths < 30 m, within 20 km from shore, temperatures > 12 C, and salinity > 25 PSU (Figure 3). However, not all surveys collected the full set of environmental variables (i.e., temperature and salinity at both bottom and surface). Therefore, to assemble as complete a dataset as possible, average monthly values from the GLORYS oceanographic model (Castillo-Trujillo et al, 2023) were used to impute environmental variables where observed data were unavailable. The accuracy of GLORYS estimates was evaluated by comparing the observed vs predicted values for survey tows where empirical data were available. Bottom and surface temperatures were the only variables where GLORYS estimates were reasonably accurate (Pearson correlation R = 0.94 and 0.79, respectively) and were used to impute missing data (18% missing for surface temperature; 7% missing for bottom temperature)(Figure 4). Empirical salinity values were unavailable for approximately 25% of tows and GLORYS salinity estimates were a poor predictor of observed values (R = -0.51). Therefore, salinity was not considered as a candidate variable during model development.

Model Development

The R package *sdmTMB* was used to construct species distribution models that accounted for spatial pattern via Gaussian Markov random fields. This required first constructing a triangular model mesh (n knots = 170) that encompassed the spatial extent of all survey observations (Figure 5). Preliminary model fits revealed that a delta-lognormal error structure provided a better representation of the data than either a negative binomial or tweedie distribution. Using catch weight per tow (kg) as the response variable, increasingly complex candidate models were attempted, using a forward selection of candidate variables via the Bayesian Information Criterion (BIC). After the optimal set of predictor variables was selected, model variants were attempted that also allowed the spatial pattern to vary seasonally (i.e., by quarter or bi-monthly) and/or over time (i.e., by year, with either IID, AR1 or random walk correlation structure).

Population Biomass Estimates

Continuous rasters of depth (GEBCO digital elevation model) and monthly average bottom temperature (GLORYS oceanographic model) were used to inform SDM predictions of anchovy CPUE across the entire model domain at 5km x 5km resolution in the units of the NEFSC-Albatross survey, for each month and year, 1985-2023. GLORYS data are only available from 1993-present, so the average of monthly rasters from the five earliest available years (1993-1997) were used to represent the monthly bottom temperature for 1985-1992. Predicted CPUE rasters were then resampled to a higher spatial resolution (1 km x 1 km), which enabled a more accurate masking of the shoreline. CPUE (Kg/tow) was translated to biomass density (kg/km^2) by dividing predicted values by the expected area swept by the NEFSC-Albatross trawl (Figure 7). Expected area swept by the NEFSC-Albatross trawl was estimated by the product of the average tow length (3.175 km) and the expected net width as a function of depth (Figure 8).

The observed biomass density from a bottom trawl survey is often a small fraction of the true biomass density, particularly for small pelagic fish such as anchovies. A “catchability” scalar (Q) can be estimated by comparing trawl-based estimates to stock assessment estimates of population biomass. Unfortunately, this was not possible for anchovies because they are not an assessed species. However, several studies have made rigorous attempts to estimate the biomass of anchovies within Chesapeake Bay (Luo and Brandt 1993; Wang and Houde 1995; Rillings and Houde 1999; Jung and Houde 2004). These estimates were converted to biomass density (kg/km^2) and compared to the SDM predictions for equivalent years, seasons, and areas. For each published estimate, a Q value was calculated by dividing the SDM density by the literature value (Table 4). The SDM predicted rasters of biomass density (kg/km^2 in NEFSC units) were then divided by the average of these Q values (0.0443) to achieve population-scale estimates of biomass density. The resulting values were multiplied by the areal extent of each pixel (1 km^2) and summed across the entire domain to obtain coastwide estimates of anchovy biomass.

The seasonal profile of total anchovy biomass was calculated by averaging the coastwide estimates across all years, by month. Likewise, the timeseries of annual biomass was calculated by averaging the coastwide estimates across all months, by year.

Results

The model form that achieved the lowest BIC included a survey effect (with NEFSC split into separate Albatross and Bigelow series), a non-linear depth effect that varied seasonally, a non-linear bottom temperature effect, and a circular seasonal effect. In addition, the best fitting model also had a spatial pattern that varied seasonally (by 2-month block), as well as over time (according to an AR1 process)(Table 3; Figure 6).

Total coastwide anchovy biomass estimated by the SDM had two seasonal peaks (April and November), with the fall peak being slightly larger than the spring peak (Figure 9). The predicted pattern within Chesapeake Bay was somewhat different, with peaks in May and October, and very low levels in January-February. This seasonal pattern is very similar to that reported for a 13-year dataset of year-round trawl observations in Chesapeake Bay (Horwitz, 1987; Houde and Zastrow, 1991), and to the findings of Wang and Houde (1995).

Average annual coastwide biomass of anchovy predicted by the SDM generally varied between 200,000 and 800,000 mt over the period 1985-2023. There were several standout years of either high (1995, 2000, 2012, 2017) or low (2003, 2005, 2021, 2022) biomass with 2012 being the highest (829.9 mt) and 2005 being the lowest (204.2 mt)

References

Castillo-Trujillo, A.C., Kwon, Y.O., Fratantoni, P., Chen, K., Seo, H., Alexander, M.A. and Saba, V.S., 2023. An evaluation of eight global ocean reanalyses for the Northeast US continental shelf. *Progress in Oceanography*, *219*, p.103126.

GEBCO Bathymetric Compilation Group: GEBCO 2023 Grid – a continuous terrain model of the global oceans and land, NERC EDS British Oceanographic Data Centre NOC [data set], https://doi.org/10.5285/f98b053b-0cbc-6c23-e053-6c86abc0af7b, 2023.

Houde and Zastrow. 1991. Bay anchovy Anchoa mitchilli. University of Maryland, Center for Estuarine and Environmental Studies. pp.1-14.

Horwitz, R.J., 1987. Fish. In: Heck, K.L (ed). *Ecological Studies in the Middle Reach of Chesapeake Bay*, Calvert Cliffs. Spring-Verlag. P. 167-224

Jung, S. and Houde, E.D., 2004. Recruitment and spawning-stock biomass distribution of bay anchovy (Anchoa mitchilli) in Chesapeake Bay. *Fishery Bulletin*, *102*(1), pp.63-78.

Luo, J. and Brandt, S.B., 1993. Bay anchovy Anchoa mitchilli production and consumption in mid-Chesapeake Bay based on a bioenergetics model and acoustic measures of fish abundance. *Marine Ecology-Progress Series*, *98*, pp.223-223.

Rilling, G.C. and Houde, E.D., 1999. Regional and temporal variability in distribution and abundance of bay anchovy (Anchoa mitchilli) eggs, larvae, and adult biomass in the Chesapeake Bay. *Estuaries*, *22*, pp.1096-1109.

Wang, S.B. and Houde, E.D., 1995. Distribution, relative abundance, biomass and production of bay anchovy Anchoa mitchilli in the Chesapeake Bay. *Marine Ecology Progress Series*, *121*, pp.27-38.

Tables and Figures

Table 1. Trawl survey datasets used to construct the SDM, with the number of tows and % with anchovies, for the period 1985-2023.

Acronymn	Survey	n Tows	% Positive
NEFSC	Northeast Fisheries Science Center	22,833	10%
NEAMAP	Northeast Area Monitoring and Assessment Program	4,724	51%
MADMF	Massachusetts Division of Marine Fisheries	7,358	5%
CT-LISTS	Connecticut Long Island Sound Trawl Survey	4,122	28%
NJOT	New Jersey Ocean Trawl Survey	5,813	39%
DBTrawl	Delaware Bay Trawl Survey	1,863	95%
ChesMMAP	Chesapeake Bay Multispecies Monitoring and Assessment Program	7,440	22%
NC-Pamlico	North Carolina Pamlico Sound Trawl Survey	4,966	63%
	Total	59,119	25%

Table 2. Candidate variables used in model selection.

Variable	Description
SURV	survey (factor)
SURV_SPLIT	same as SURV, but with NEFSC split into NEFSC_ALB, NEFSC_BIG
YEAR	year (numeric)
(1|YFAC)	random effect for year
s(DEPTH)	smoother of depth, in meters
s(DSHORE)	smoother of distance from shore, in km
s(BOTTEMP)	smoother of bottom temperature, in C
s(SURFTEMP)	smoother of surface temperature, in C
s(J)	circular smoother of day-of-year
s(DEPTH,J)	smoothed interaction of DEPTH x J (circular)
s(DSHORE,J)	smoothed interaction of DSHORE x J (circular)
s(BOTTEMP,J)	smoothed interaction of BOTTEMP x J (circular)
s(SURFTEMP,J)	smoothed interaction of SURFTEMP x J (circular)
QFAC	quarter (Jan-Mar; Apr-Jun; Jul-Sep; Oct-Dec)
BFAC	bi-month (Jan-Feb; Mar-Apr; May-Jun; Jul-Aug; Sep-Oct; Nov-Dec)

Table 3. Summary of stepwise process for model selection. Each step represents an increment in model complexity that achieved the greatest reduction in BIC. The model listed in bold (step 7) was selected as the best fitting model.

Step	Model	Spatio-temporal	Spatial-varying	EDF	BIC	ΔBIC
1	~ 1			7	68061	*NA*
2	~ s(DEPTH,J)			9	63288	-4773
3	~ s(DEPTH,J) + SURV_SPLIT			25	61168	-2120
4	~ s(DEPTH,J) + SURV_SPLIT + s(BOTTEMP)			29	60402	-766
5	~ s(DEPTH,J) + SURV_SPLIT + s(BOTTEMP) + s(J)			31	60216	-186
6	~ s(DEPTH,J) + SURV_SPLIT + s(BOTTEMP) + s(J)	YEAR (ar1)		35	55668	-4548
7	**~ s(DEPTH,J) + SURV_SPLIT + s(BOTTEMP) + s(J)**	**YEAR (ar1)**	**0 + BFAC**	**45**	**54238**	**-1430**

Table 4. Comparison of literature estimates of anchovy biomass in Chesapeake Bay to equivalent estimates from the SDM, predicted in kg/km^2 for the NEFSC-Albatross survey. The "catchability" scalar (Q) was calculated by dividing the SDM estimate by the literature value.

				Density (Kg / Km^2)		
Source	Years	Season	Area	Literature	SDM	Q
Luo and Brandt, 1993	1989	Sep	Mid CB	10,700	316.4	0.0296
Luo and Brandt, 1993	1989	Nov	Mid CB	14,700	260.4	0.0177
Wang and Houde, 1995	1990-1991	Apr	Upper + Mid CB	9,503	118.0	0.0124
Wang and Houde, 1995	1990-1991	Jun	Upper + Mid CB	4,924	128.9	0.0262
Wang and Houde, 1995	1990-1991	Aug	Upper + Mid CB	16,433	67.4	0.0041
Wang and Houde, 1995	1990-1991	Oct	Upper + Mid CB	29,974	347.7	0.0116
Wang and Houde, 1995	1990-1991	Dec	Upper + Mid CB	573	124.9	0.2182
Rilling and Houde, 1999	1993	Jun	All CB	4,250	129.5	0.0305
Rilling and Houde, 1999	1993	Jul	All CB	4,206	91.3	0.0217
Jung and Houde, 2004	1995	Annual	All CB	4,353	242.5	0.0557
Jung and Houde, 2004	1996	Annual	All CB	3,083	274.2	0.0889
Jung and Houde, 2004	1997	Annual	All CB	7,617	244.4	0.0321
Jung and Houde, 2004	1998	Annual	All CB	11,425	358.5	0.0314
Jung and Houde, 2004	1999	Annual	All CB	8,705	241.3	0.0277
Jung and Houde, 2004	2000	Annual	All CB	11,607	314.7	0.0271
Jung and Houde, 2004	1996	Oct	All CB	4,900	540.5	0.1103
Jung and Houde, 2004	1998	Oct	All CB	35,000	608.0	0.0174
Jung and Houde, 2004	1995-2000	Oct	All CB	18,900	660.1	0.0349
					Mean =	0.0443

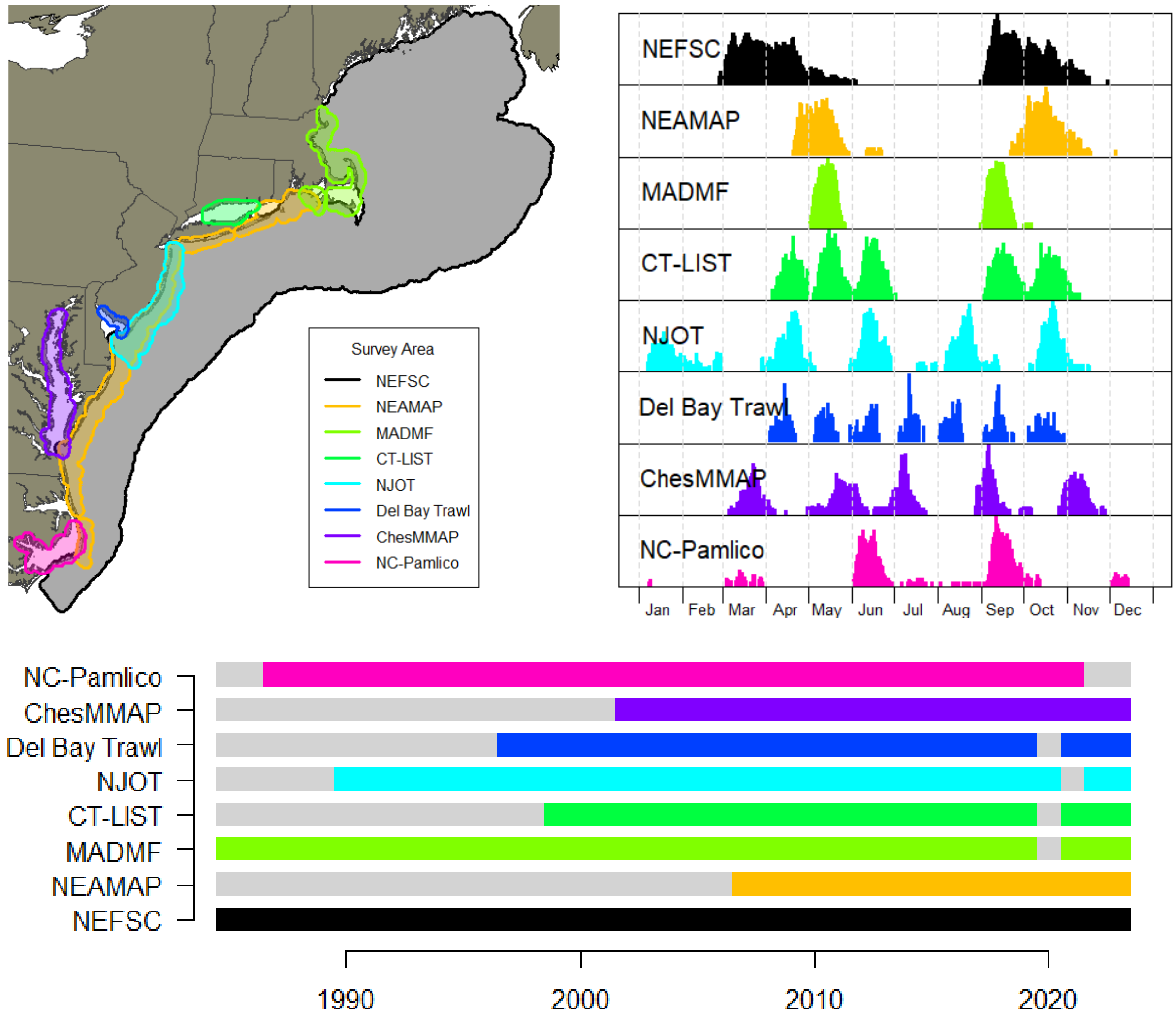

Figure 1. Spatial [top left], seasonal [top right], and temporal [bottom] extents of trawl survey data used in the SDM.

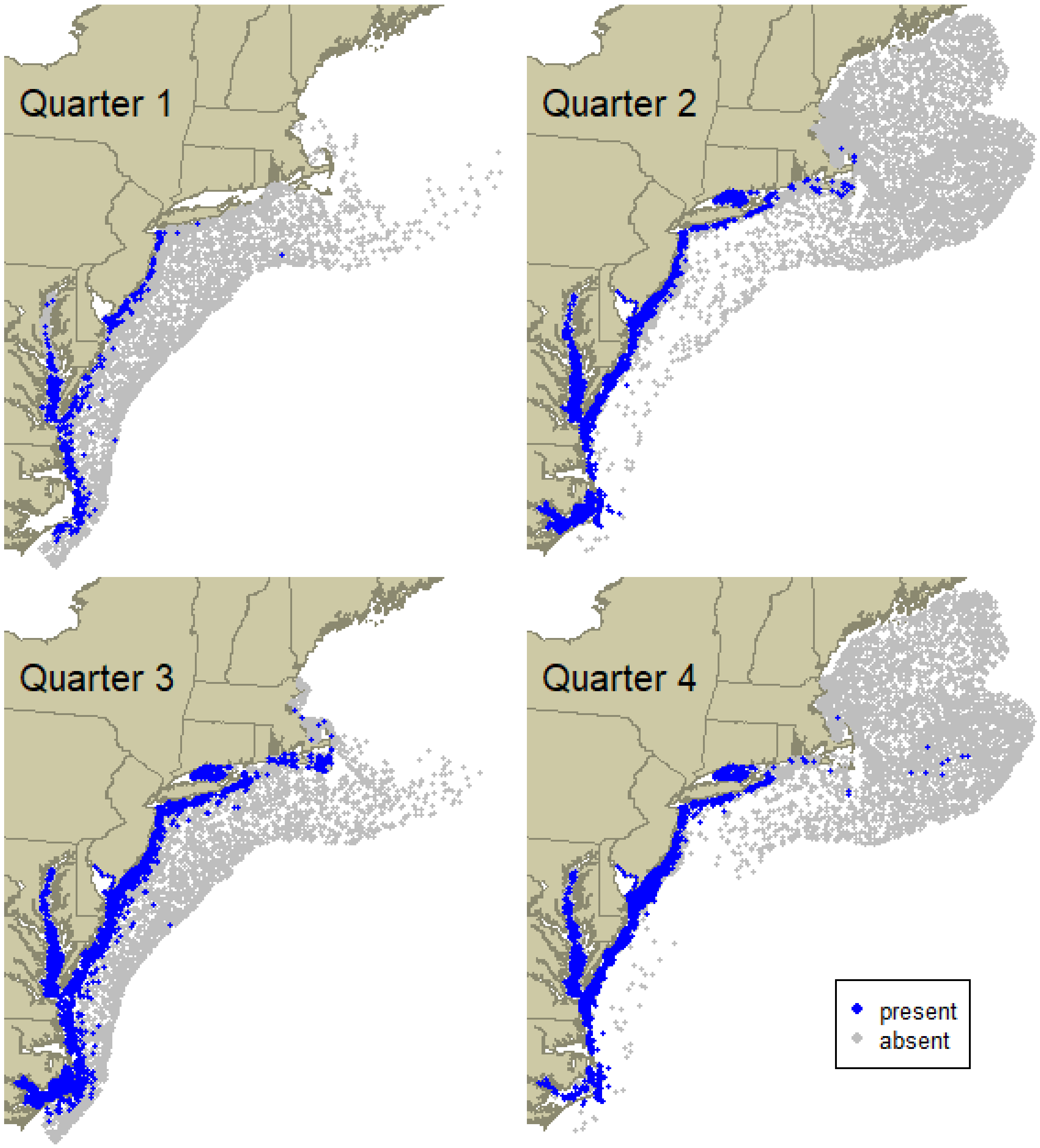

Figure 2. Observed occurrence of anchovies across all surveys, by quarter.

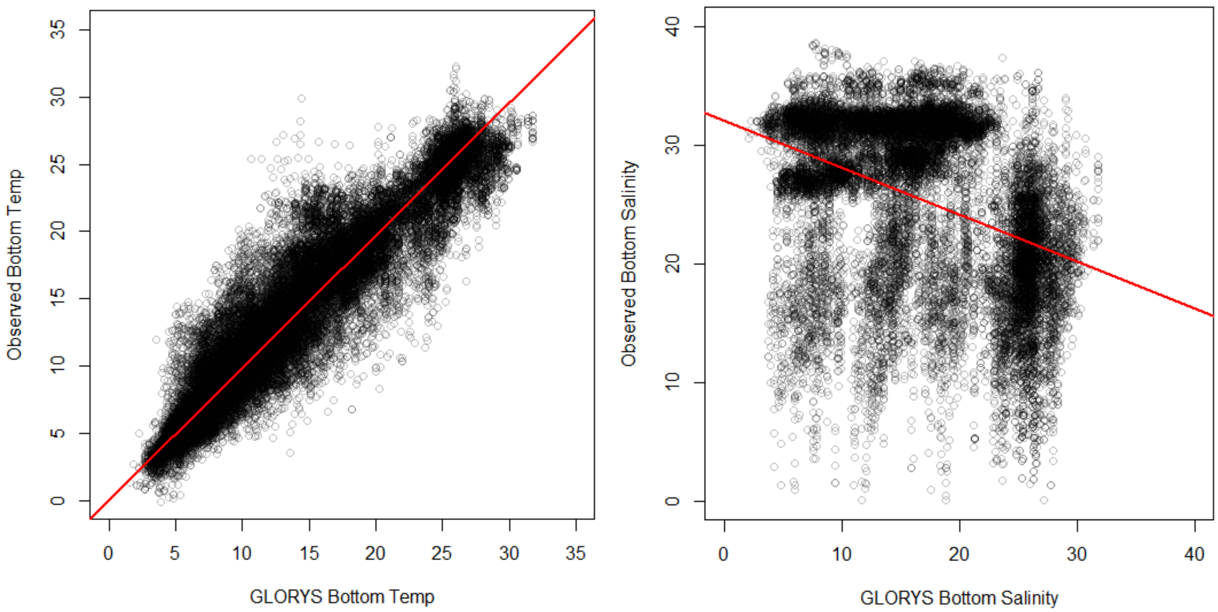

Figure 3. Observed values vs GLORYS-predicted values for bottom temperature (left) and bottom salinity (right). The solid red line represents a linear regression (OBS~PRED).

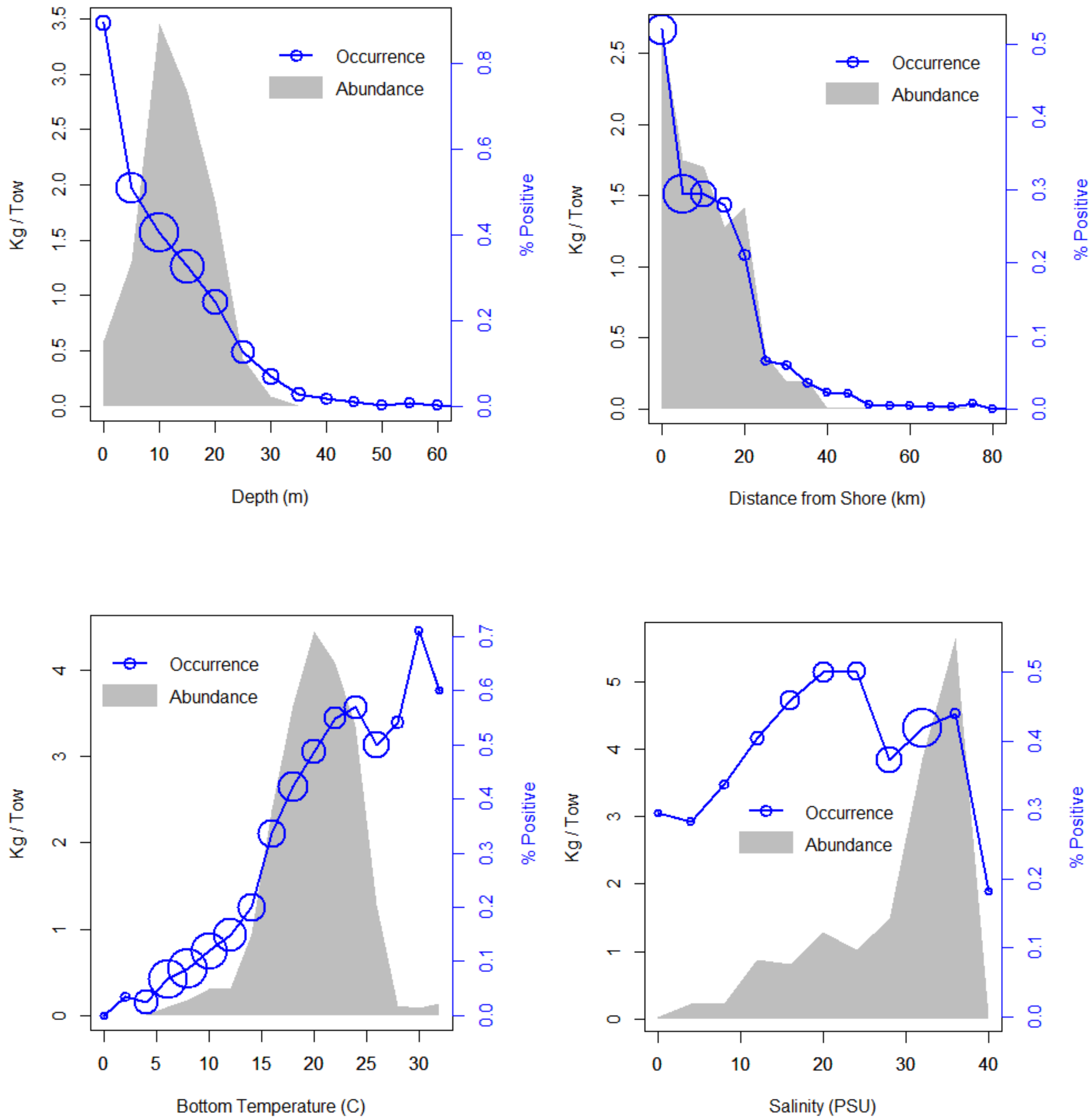

Figure 4. Observed relationship for occurrence (blue) and abundance (gray) of anchovies, with respect to depth (top left), distance from shore (top right), bottom temperature (bottom left), and bottom salinity (bottom right). The size of the bubbles is proportional to the number of observations.

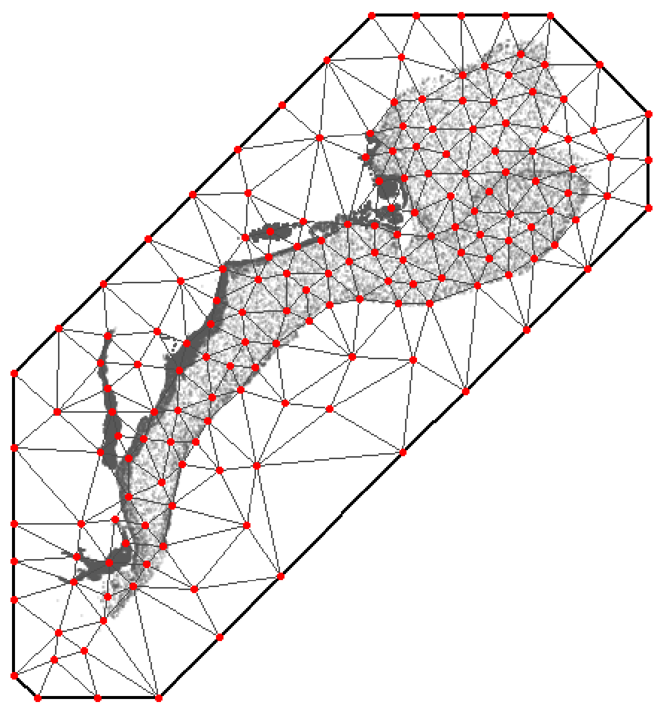

Figure 5. Triangular model mesh used to account for spatial pattern in the SDM.

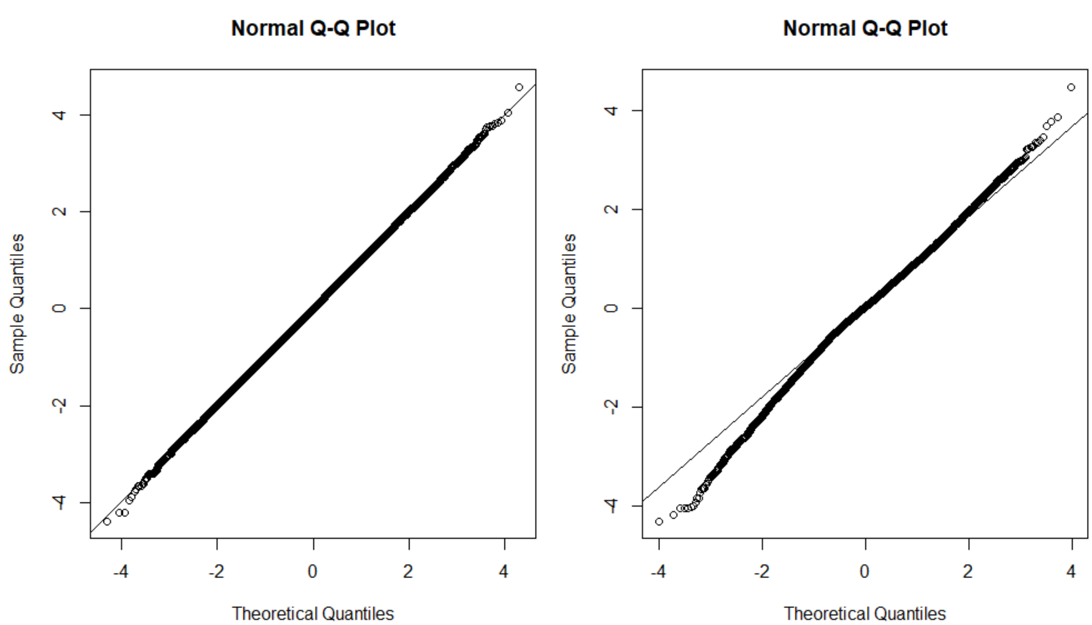

Figure 6. Fit of the selected SDM to the assumed delta-lognormal error distribution: [left] QQ-plot for the occurrence model component; [right] QQ-plot for the log(CPUE) model component.

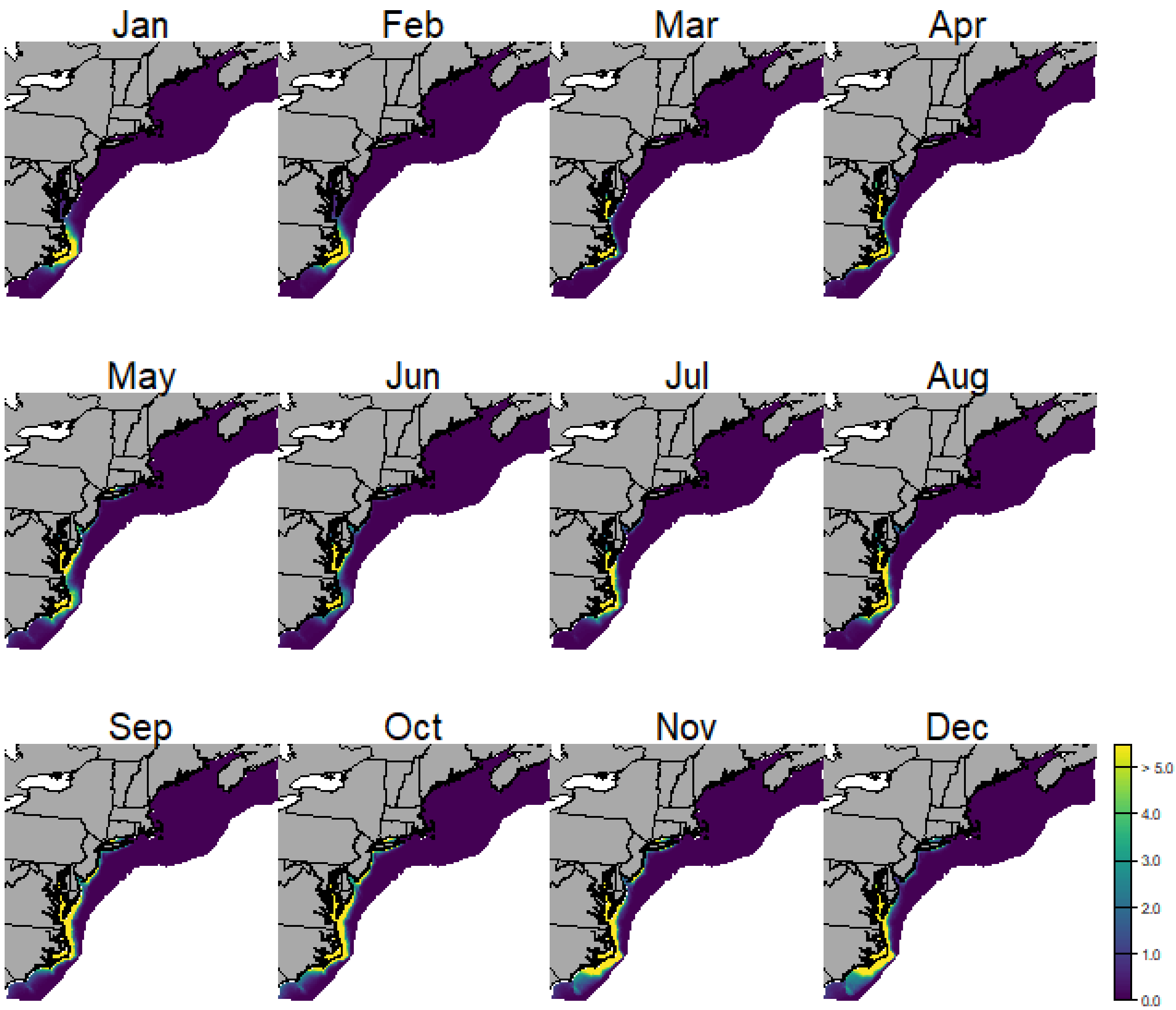

Figure 7. Average predicted distribution of anchovies by month, 1985-2023. Units are in catch per tow for the NEFSC (Albatross) survey.

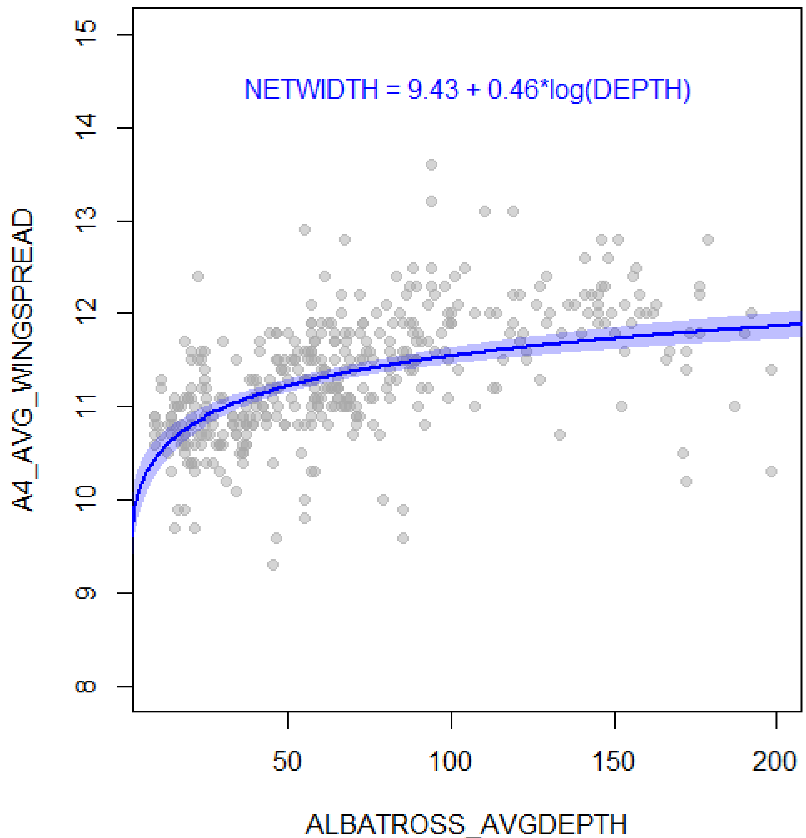

Figure 8. Observed width of the NEFSC-Albatross net, as a function of depth. The blue line represents a fitted model of the form: NETWIDTH ~ log(DEPTH).

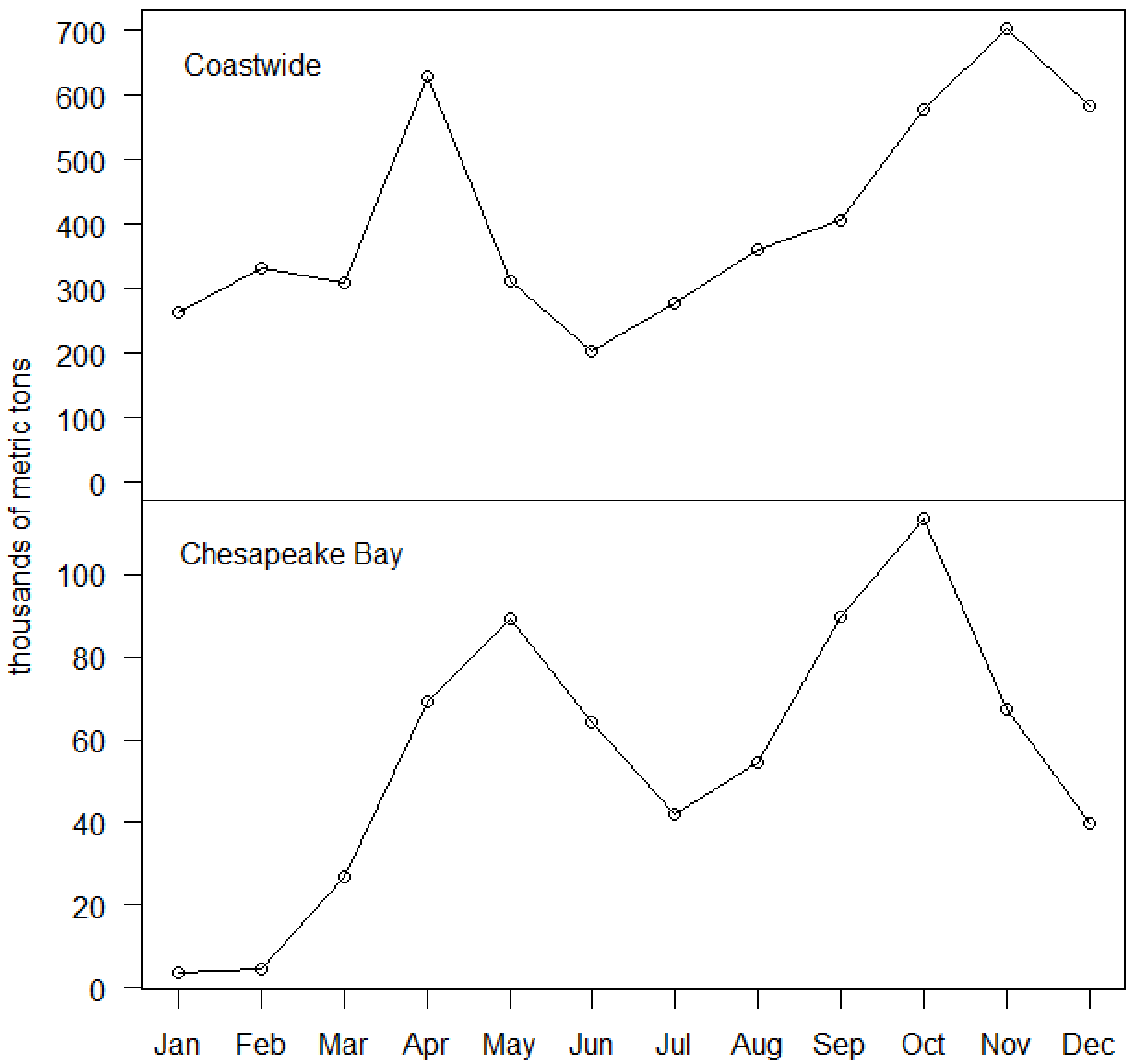

Figure 9. Seasonal profile of average anchovy biomass across the entire NWACS model domain (top) and in Chesapeake Bay (bottom), 1985-2023.

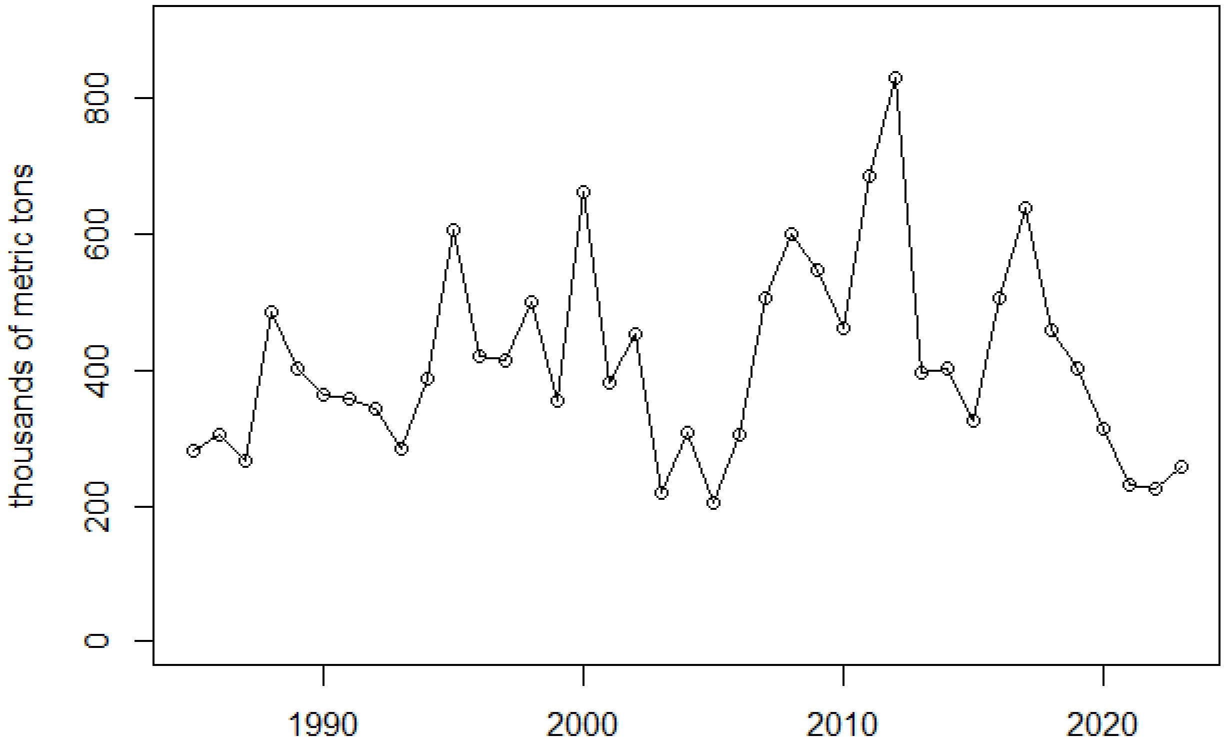

Figure 10. Trajectory of average annual anchovy biomass within the NWACS model domain.