A predictive krill distribution model for *Euphausia pacific*a and *Thysanoessa spinifera* using scaled acoustic backscatter in the Northern California Current

Derville, S.[1,2], Fisher, J.L.[3], Kaplan, R.[4], Bernard, K.S.[4], Phillips, E. M.[5], Torres, L.G.[1]

[1] Geospatial Ecology of Marine Megafauna Lab, Marine Mammal Institute, Department of Fisheries, Wildlife and Conservation Sciences, Oregon State University, Newport, OR, United States
[2] UMR ENTROPIE (UR-IRD-IFREMER-CNRS-UNC), Nouméa, New Caledonia
[3] Fish Ecology Division, Northwest Fisheries Science Center, National Oceanic and Atmospheric Administration, Newport, OR, Unites States
[4] College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, Oregon, United States
[5] Fishery Resource Analysis and Monitoring Division, Northwest Fisheries Science Center, National Oceanic and Atmospheric Administration, Seattle, WA, United States

SUPPLEMENTARY MATERIAL

Table S1: Hydroacoustic and bongo net tow survey effort per season and per year. Number of distinct days with echosounder data and number of net tow samples are reported. The months of February, May, and September are the months targeted by NCC surveys but effort may occasionally extend by a few days before or after.

Surveys	2018	2019	2020	2021	2022
February	3 tow samples		4 days 6 tow samples		
May	10 days 21 tow samples	11 days 4 tow samples		3 days 10 tow samples	11 days 18 tow samples
September	8 days 18 tow samples	9 days 19 tow samples	9 days 16 tow samples		3 days 9 tow samples

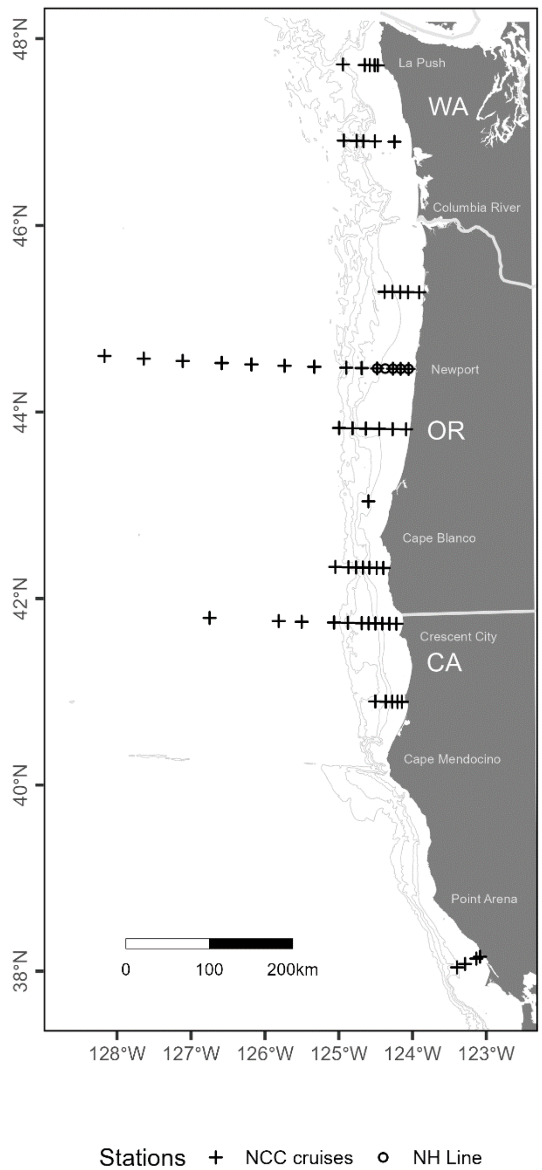

Figure S1: Map of NCC surveys and NH Line stations at which bongo net tows were performed and whose biomass data were included in this study. Land is shown in dark gray. Isobaths (200 m, 500 m, 1000 m, and 1,500 m deep) are represented with gray lines.

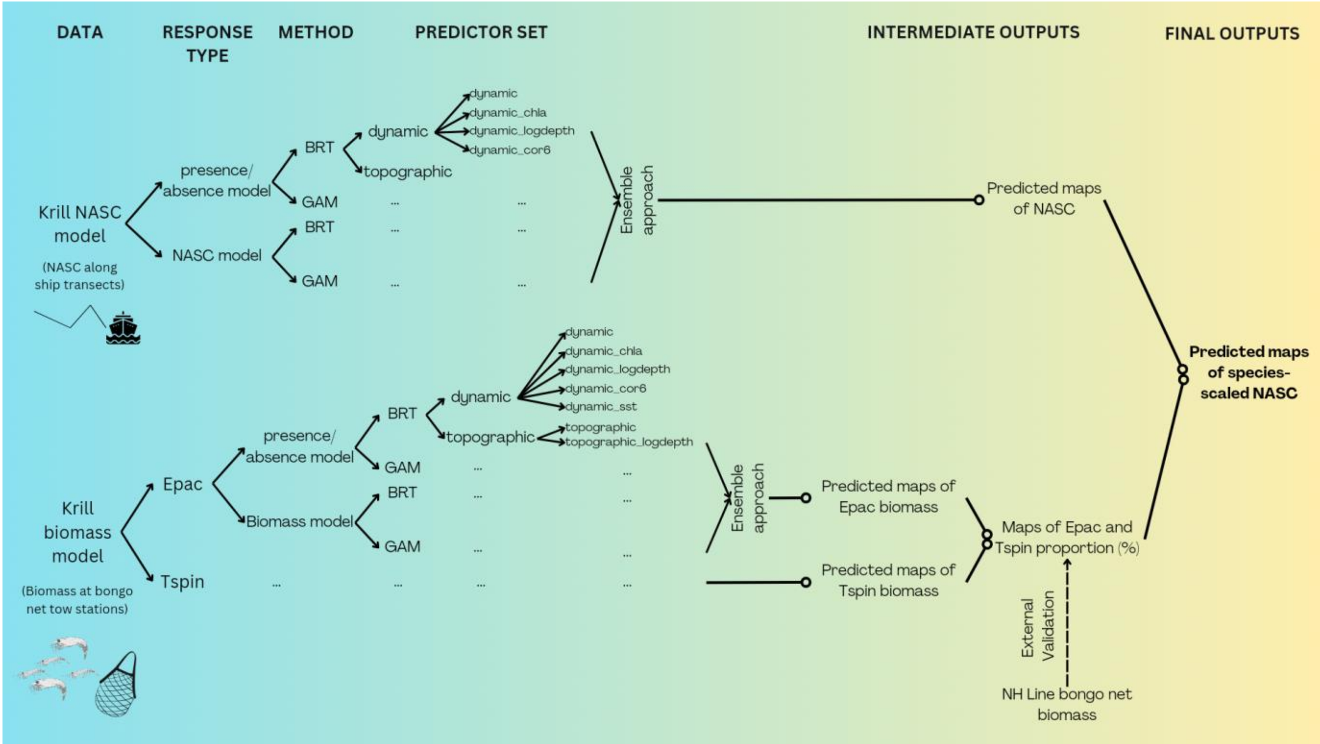

Figure S2: Schematic representation of the modelling framework. The symbols "…" indicates that the same variables were tested/methods were conducted as listed above. Species names are abbreviated as Tspin (*Thysanoessa spinifera) and (Euphausia pacifica).*

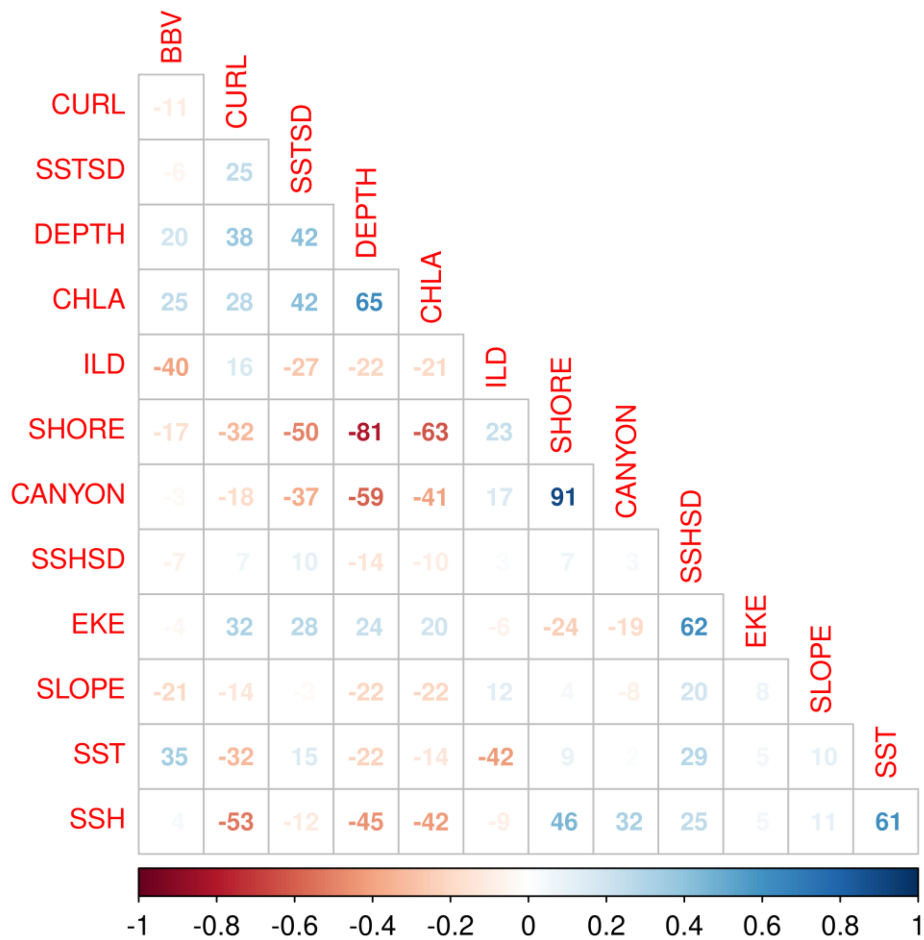

Figure S3: Pearson correlation coefficients calculated between pairs of environmental variables extracted at the time and position of the daytime NASC data. Environmental predictors are the following: distance to canyons (CANYON in km), log-transformed seabed depth (DEPTH in m), sea surface temperature (SST in °C) and its spatial standard deviation (SSTSD calculated over 0.3° squares), sea surface height (SSH in m) and its standard deviation (SSHSD calculated over 0.3° squares), log-transformed eddy kinetic energy (EKE calculated from eastward and northward surface current velocities, $kg{\cdot}m^2{\cdot}s^{-2}$), wind stress curl (CURL in $Newton.m^{-3}$), isothermal layer depth (ILD in m) and bulk buoyancy frequency (BBV in s^{-1}). DIS_SHORE is the distance to the coast. It was strongly correlated with distance to canyons and seabed depth and was therefore discarded.

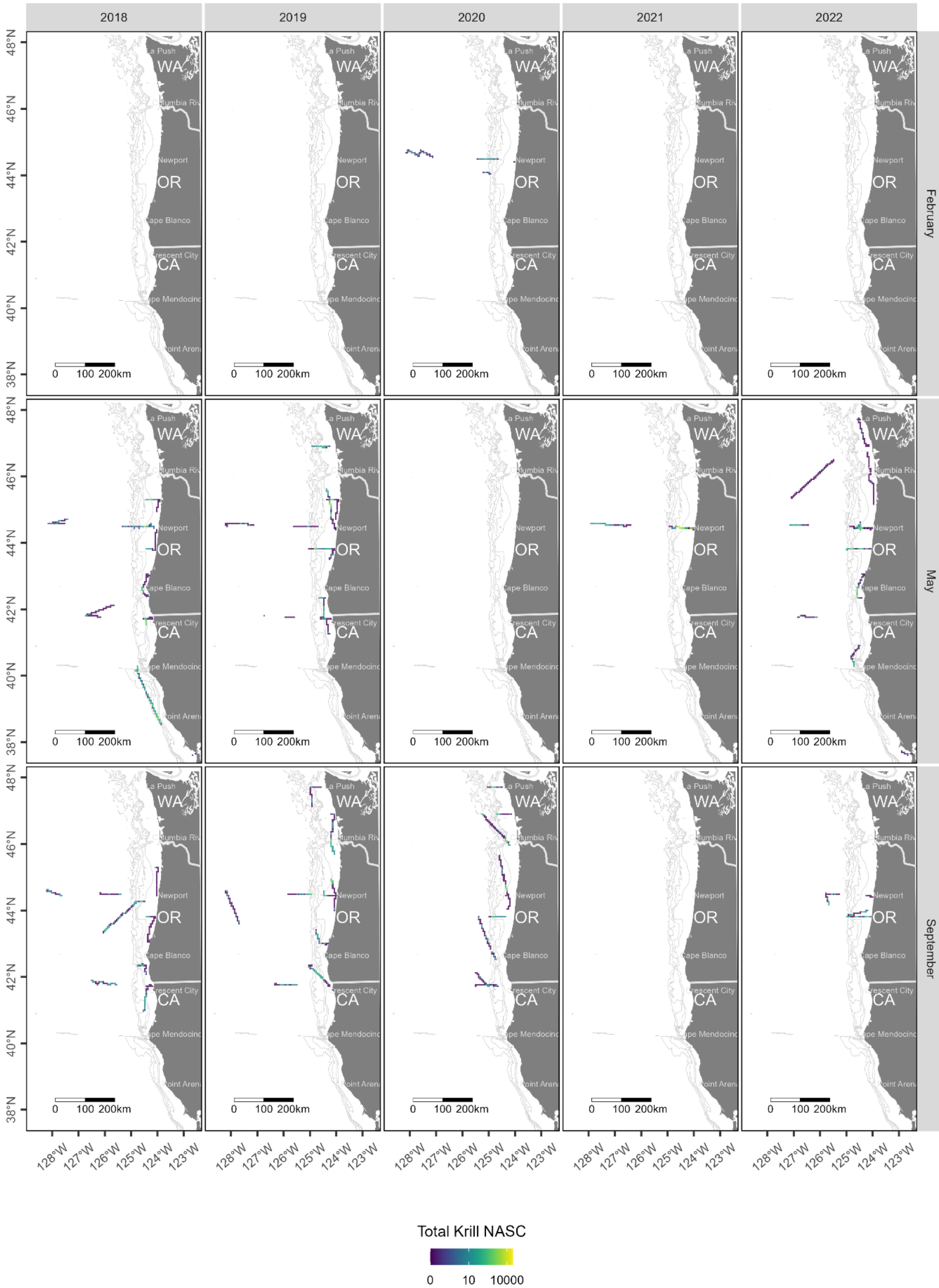

Figure S4: Maps of krill NASC calculated along transects and aggregated in 5 km resolution grids (daily data are overlayed) per season, per year. Note that several seasonal surveys did not occur or NASC data were not available, particularly in the winter (e.g., Feb 2018, 2019, 2021, and 2022). Land is shown in dark gray. Isobaths (200 m, 500 m, 1000 m, and 1,500 m deep) are represented with gray lines.

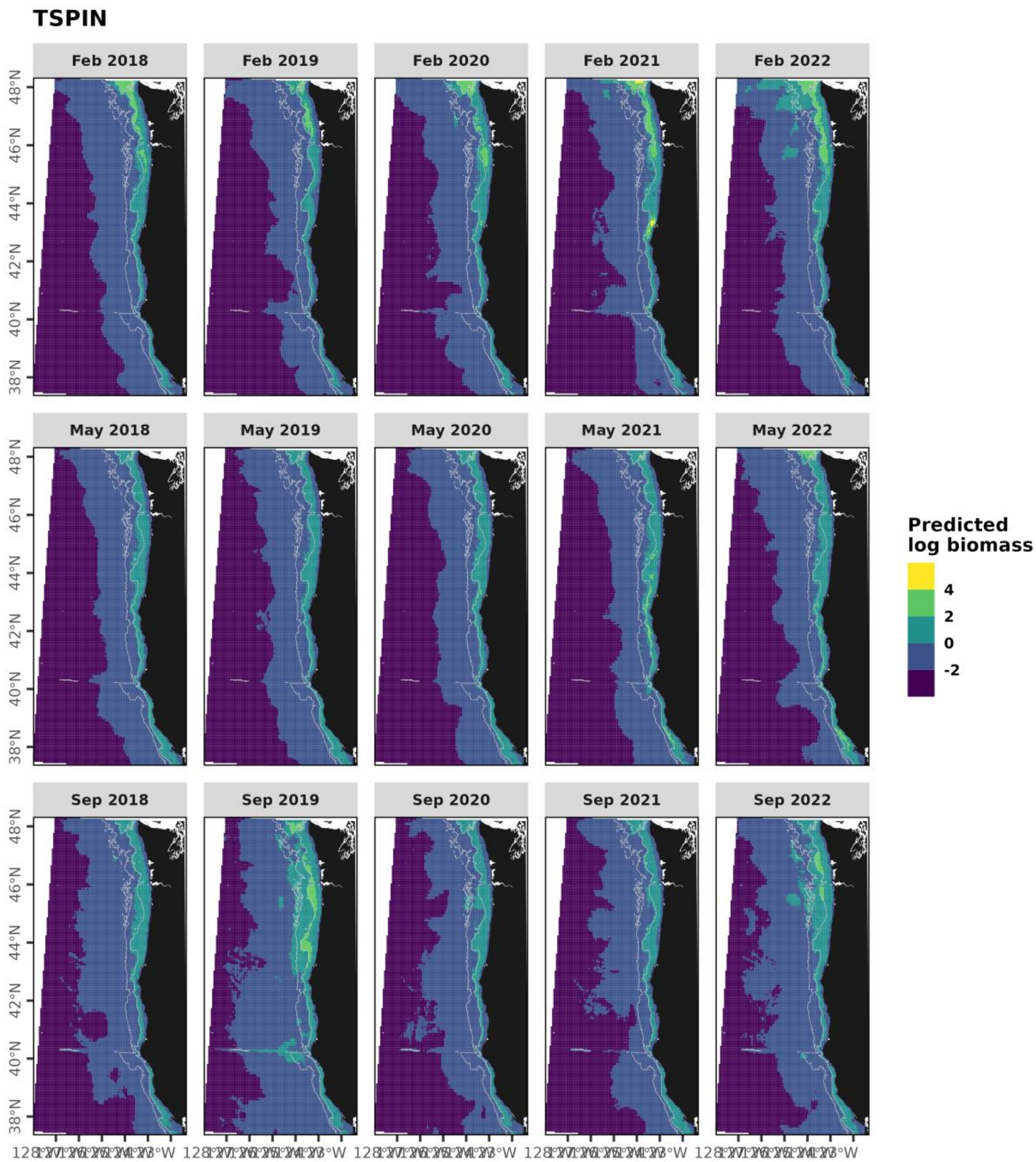

Figure S5: Predicted biomass of *Thysanoessa spinifera* generated from the dynamic_logdepth GAM models (presence-absence GAM x biomass GAM) for the months of February, May and September 2018-2022. Predictions show an unrealistic high biomass in February compared to other months of year.

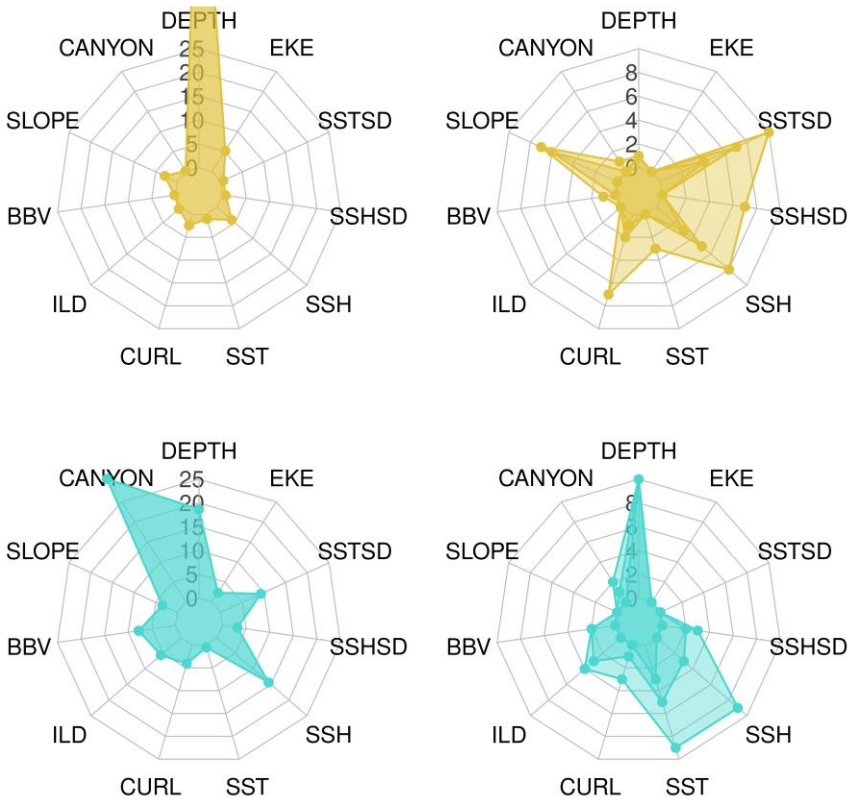

Figure S6: Predictor contributions to the dynamic_logdepth BRT presence/absence model (left) and GAM biomass (right) models of *Euphausia pacific*a (top, yellow) and *Thysanoessa spinifera* (bottom, blue). Predictor contributions are measured either by the percent of contribution estimated in the BRT presence/absence model. redictor contributions are measured either by the percent of contribution estimated in the BRT presence/absence model or by the number of cross-validation folds in which the approximate smooth significance p-values were below 0.05, 0.01 or 0.001 (shown with increasingly dark color shades) in the GAM abundance model. Environmental predictors are the following: distance to canyons (CANYON in km), log-transformed seabed depth (DEPTH in m), seabed slope (SLOPE in radians), sea surface temperature (SST in °C) and its spatial standard deviation (SSTSD calculated over 0.3° squares), sea surface height (SSH in m) and its standard deviation (SSHSD calculated over 0.3° squares), log-transformed eddy kinetic energy (EKE calculated from eastward and northward surface current velocities, kg·m^2·s^{-2}), wind stress curl (CURL in Newton.m^{-3}), isothermal layer depth (ILD in m) and bulk buoyancy frequency (BBV in s^{-1}).

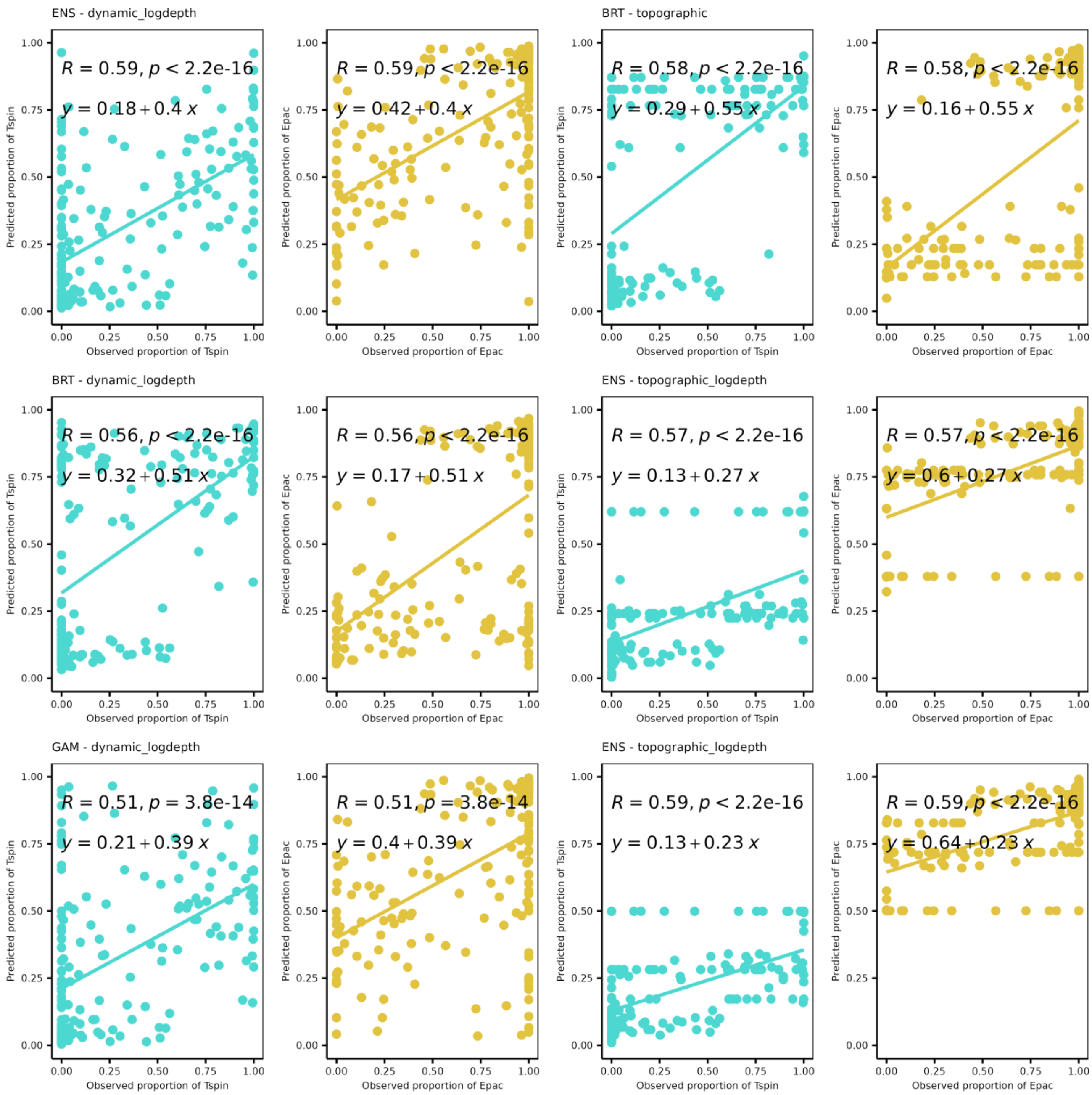

Figure S7: Predicted vs observed proportion of *Euphausia pacific*a (Epac: right, yellow) and *Thysanoessa spinifera* (Tspin: left, blue) at the NH Line stations sampled year round and the NCC stations sampled around February, May, and September, between 2018 and 2022. Each panel corresponds to a different combination of statistical algorithms (Ensemble, BRT, GAM) and predictor set (dynamic_logdepth or topographic and topographic_logdepth) to model species-specific krill presence/absence and biomass. The ensemble approach tested here combines a BRT presence/absence model and a GAM biomass model. The Pearson coefficient of correlation and associated p-value of the regression line between observed and predicted values is represented on each panel, along with the equation of the regression line. The selected approach is that of the Ensemble approach with a topographic_logdepth predictor set (bottom right panels).

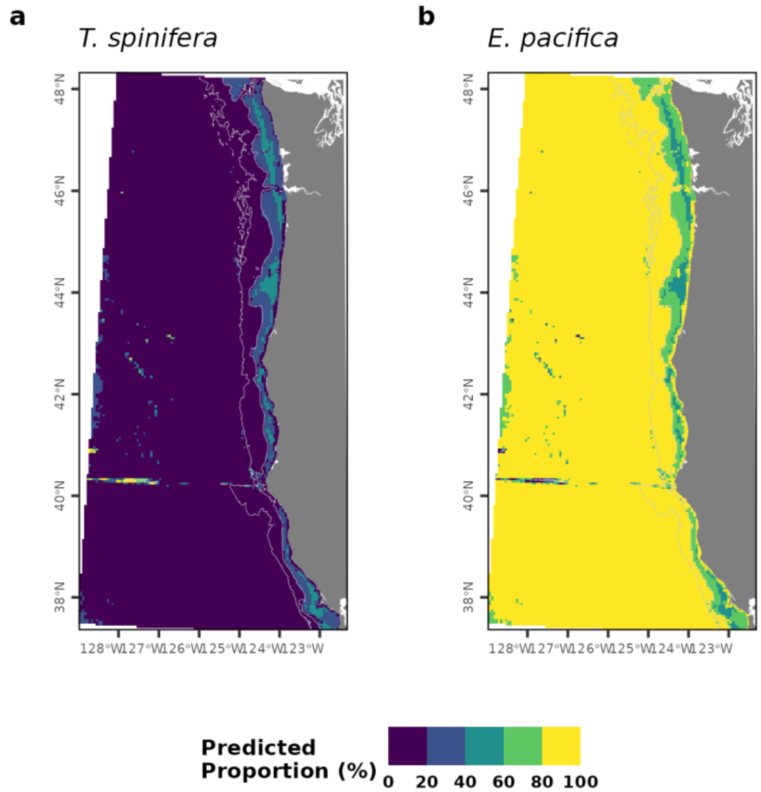

Figure S8: Predicted proportion of *Euphausia pacifica* and *Thysanoessa spinifera* biomass over the entire NCC domain. Land is shown in black. Isobaths (200 m, 500 m, 1000 m, and 1,500 m deep) are represented with gray lines.

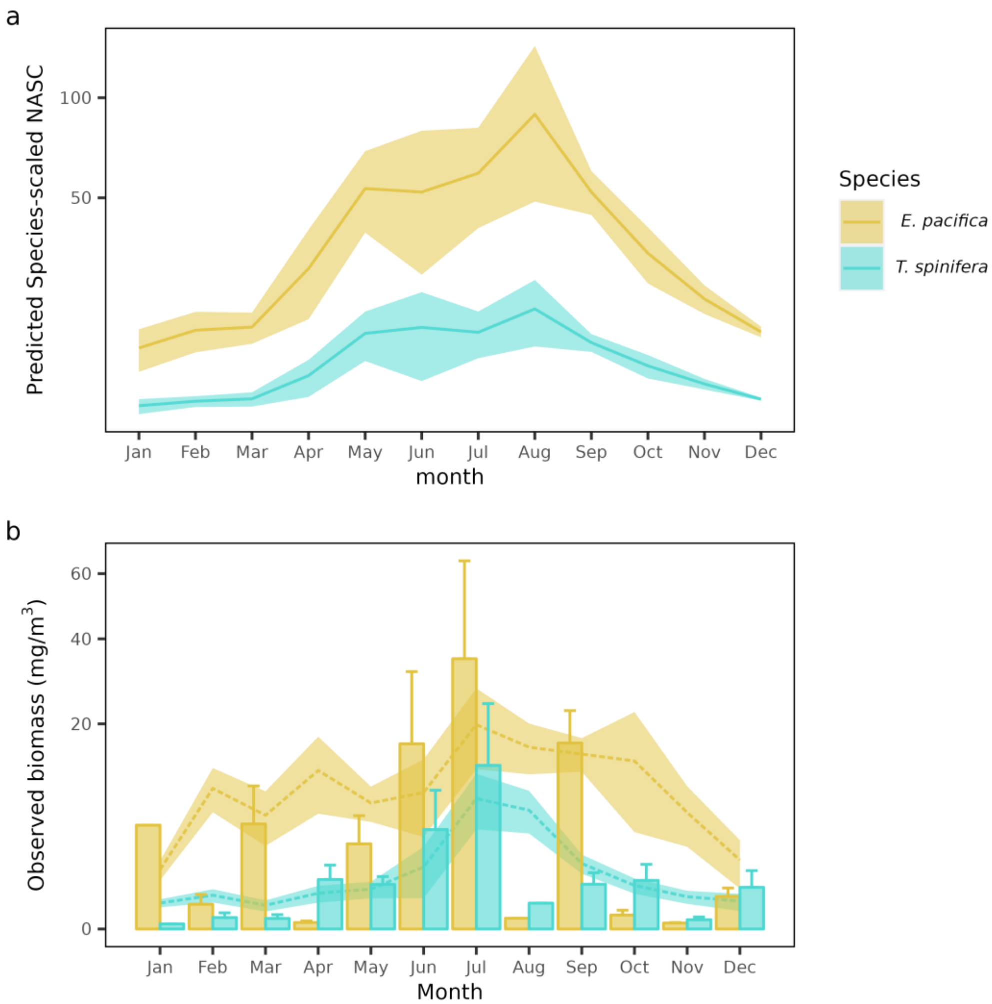

Figure S9: Predicted (a) vs observed (b) patterns of krill species-scaled NASC and biomass at the latitude of the NH Line in relation to month of the year. Panel a shows the summed predicted species-scaled NASC per species, averaged by year x month and then by month. Colored ribbons represent the standard deviation across years. Panel b shows the mean observed biomass per species per month, averaged across five stations (NH05, NH10, NH15, NH20, and NH25). Bars represent the observed biomass across five years (2018-2022), while dashed lines represent the observed biomass across 21 years (2001-2022). Error bars and colored ribbons represent the standard error of the average across years by month. Note that between 2018 and 2022, the month of August was only sampled once along the NH Line in 2019. August bars are therefore not as representative as other summer months such as July where both the 2018-2022 (bars) and 2001-2022 (dashed line) observed biomass are consistently high. The y-axis is square-root transformed.