

Incorporating spatial distribution at age improves predictive ability of SDMs

Abstract

Species distribution models (SDMs) are widely used in fisheries and ecology, providing insights into species' spatial distributions. While many SDMs incorporate external information, such as environmental or habitat features, the majority overlook age-specific information, despite the known influence of age structure on fish distribution and abundance. Including age information may be particularly valuable for species exhibiting age-related spatial patterns driven by ontogenetic shifts, recruitment dynamics, and selective fishing pressures. Here, we develop an age-structured SDM framework using data from the west coast of the USA, and concentrate on two species with distinct life histories: North Pacific hake (*Merluccius productus*) and sablefish (*Anoplopoma fimbria*). Our framework incorporates spatiotemporal models to predict age-specific distributions. We validate our approach using the recursive nature of age data to forecast the spatial distribution of fish by age class across survey locations in subsequent years; these results highlight that forecast models have predictive skill (more so for sablefish than hake) and the predictive ability is highest for older age classes. Output from our models can be used to predict the spatial age composition, which may be particularly useful for large cohorts (hake from 2010 and 2014 cohorts; sablefish from 2013, 2016, and 2020 cohorts). Our approach is applicable to a wide range of survey types and platforms around the world, and supports the development of spatially explicit management strategies and optimal allocation of fishing effort.

Introduction

Species distribution models (SDMs) are widely used in marine ecology to understand and predict the spatial distribution of species based on environmental and demographic variables (Melo-Merino, Reyes-Bonilla, and Lira-Noriega 2020). These models have informed management decisions for conservation and resource allocation, especially in the face of changing environmental conditions and human impacts, such as fishing pressure (Guisan and Thuiller

2005; Elith and Leathwick 2009). Recent advances in SDM applications in the marine environment have incorporated external drivers such as temperature (Fredston et al. 2021), habitat (Phillips et al. 2017; Laman et al. 2018). A second use of these models has been to predict temporal trends in species biomass from spatiotemporal fisheries independent survey data (Thorson et al. 2015). Trends estimated from SDMs may be included in integrated population models (stock assessments) for commercially fished species, or used to identify risks to populations of conservation concern. Recent developments in these models have included depth (Johnson, Thorson, and Punt 2019) and habitat variables (Cao et al. 2017) to improve trend estimation.

While many SDMs around the world have included variables such as temperature, depth, or habitat, they have generally overlooked age structure, and instead modeled either species occurrences or densities across all age or size classes. Omitting information on age or size may be partially dependent on data availability (ageing fish generally requires sampling otoliths, which increases monetary costs of sampling programs), but may also be driven by the modeling choices of analysts. Much of the previous research around the world has demonstrated that age structure is a fundamental characteristic of fish populations, influencing spatial distribution and abundance due to mechanisms like ontogenetic shifts, recruitment variability, and selective fishing pressures (Rijnsdorp et al. 2018). Ignoring age-specific information in SDMs may lead to incomplete or inaccurate predictions, especially for species with strong age-related spatial gradients.

Incorporating age structure into SDMs may provide more accurate predictions by capturing spatial and temporal variability related to life history traits, age-specific habitat preferences, and population dynamics. Fish often exhibit ontogenetic shifts, where younger individuals occupy different habitats than adults due to factors like predator avoidance, habitat preferences, and diet changes (Werner and Gilliam 1984; Ciannelli et al. 2004). These shifts can create age-specific distributions across habitats, particularly for species that undergo distinct life history phases. Additionally, recruitment pulses and migration patterns vary by age, influencing the spatial distribution of younger versus older fish. Fishing selectivity further accentuates these patterns by disproportionately removing older age classes, which alters the population structure and potentially shifts spatial distributions (Hilborn and Walters 1992; Methot and Wetzel 2013). By incorporating age, SDMs could better predict these dynamics and improve our understanding of species-environment relationships.

Because of the recursive nature of population dynamics, e.g. numbers at age a at time t can be used to predict numbers at age $a+1$ at time $t+1$, age-structured SDMs with predictive skill also provide a tool for forward prediction, with considerable implications for marine resource management. Reliable predictions of species distributions can help managers anticipate and mitigate issues such as bycatch risk by identifying regions with high densities of specific age cohorts (Lewison, Freeman, and Crowder 2004). Forecasting the spatial distribution of certain age classes can inform spatially explicit management strategies and help optimize fishing effort to target sustainable cohorts. For example, knowing the expected future locations of specific age classes could aid in creating time-area closures that protect vulnerable life stages or ensure

harvests focus on more resilient portions of the population. However, building robust age-structured SDMs requires high-quality, long-term datasets with consistent sampling methods across both age and space to ensure the validity of these predictions.

Around the world, age and life history information is routinely collected by a number of sampling platforms, including from fisheries and fisheries independent surveys. While fisheries dependent data may be used to inform age structured models, sampling from fisheries independent surveys is typically more consistent in space and time. The primary objective of our paper is to explore the utility of developing age - structured SDMs from fisheries independent data. Focusing on two species with contrasting life history characteristics in the North Pacific, the semi-pelagic North Pacific hake (*Merluccius productus*) and longer lived benthic sablefish (*Anoplopoma fimbria*), we first develop a flexible framework for modeling numbers at age. Second, we evaluate the ability of this framework to offer predictive skill in future forecasts, and quantify how the predictive relationship changes as a function of age. Finally, we demonstrate how model output may be used to examine changing species compositions through time, make maps of potential bycatch hotspots, and identify areas with higher concentrations of mature fish as potential targets of fishing effort.

Methods

Data

On the west coast of the USA, the West Coast Groundfish Bottom Trawl Survey (WCG-BTS) has been an annual survey conducted from 2003 - present. The WCG-BTS is designed to estimate the abundance, size, and age composition of groundfish species important to commercial and recreational fisheries found in near-bottom habitats on the west coast of the USA (Keller et al., 2017). The survey effort is concentrated in summer months and has been conducted annually since 2003 (here we use the data through 2018; data are publicly available at <https://www.nwfsc.noaa.gov/data>). Importantly, the random stratified sampling design, spatial and seasonal coverage, effort, and gears have remained relatively constant within the period we analyse. Like many other surveys around the world, no WCG-BTS survey occurred in 2020 because of the Covid-19 pandemic. Though the WCG-BTS survey samples hundreds of species, we concentrated our analysis on two of the well sampled species from the WCG-BTS, North Pacific hake and sablefish. [more on biological gradients] While otoliths are sampled continuously during the WCG-BTS survey for a broad range of species, otoliths are generally aged when species are being prioritized for stock assessment by the Pacific Fishery Management Council (PFMC); full assessments between species may occur irregularly and be sporadically updated every 5 – 10 years. Hake and sablefish represent exceptions, as the hake stock is assessed annually by co-managers from the USA and Canada (Grandin et al. 2024) and sablefish is frequently assessed because of its high commercial value. For our analysis, we used hake collected 2007 – 2019 (mean 650.7 individuals sampled per year) and sablefish collected 2003 – 2023 (mean 1314.4 individuals sampled per year).

Spatial age models

For each species, we filtered by sex to focus on females and truncated ages to focus on those with the highest data availability; for hake this included ages 1 – 5, and for sablefish ages 0 – 9 (data from these ages represented 54% and 75% of the total aged fish, respectively). For each species - age combination, we constructed a unique spatiotemporal model fit to all years except the last available year for each species. Individual fish were aggregated at the haul level, allowing us to model counts at age. We initially explored modeling counts with a Poisson or Negative Binomial distribution, but as the distributions of numbers at age are dominated by 0s, these models failed to converge. Instead, we summarized the total number of fish sampled and number of fish sampled of the focal age; this allowed counts to be modeled as a binomial family (logit link). Second, we constructed a spatiotemporal model as an extension of Generalized Linear Mixed Models (GLMMs) such that the prediction in location s and time t can be written as

$$\text{logit}(p_{s,t}) = \beta_t + \omega_s + \delta_{s,t}$$

where β_t represent time - varying intercepts modeled as a random walk $\beta_t \sim N(\beta_{t-1}, \sigma_\beta)$, the spatial field $\omega_s \sim \text{MVNormal}(0, \Sigma_\omega)$ and the spatiotemporal fields $\delta_{s,t}$ are modeled as an AR(1) process $\delta_t = \rho\delta_{t-1} + \sqrt{1 - \rho^2}\epsilon_t$, where $\epsilon_t \sim \text{MVNormal}(0, \Sigma_\epsilon)$

Spatial and spatiotemporal random fields were constructed as Gaussian Markov random fields (GMRFs) using the stochastic partial differential equation approach (SPDE) (Lindgren, Rue, and Lindström 2011; Lindgren and Rue 2015). The SPDE method models the spatial correlation between points as a Matérn covariance function with smoothness parameter $\nu = 1$. Spatial meshes for all ages and species were constructed with a cutoff distance of 50km (this distance controls the spacing of mesh vertices). Parameter estimation was done using the sdmTMB software package (Anderson et al. 2024) with R 4.3.1 (R Core Team 2024). The sdmTMB package relies on Template Model Builder (TMB) (Kristensen et al. 2016) to quickly and efficiently maximize the marginal log likelihood using auto-differentiation and the Laplace approximation to integrate out random effects. Models were evaluated for convergence (positive-definite Hessian matrix, and a maximum absolute log likelihood gradient < 0.001) and residuals diagnostics were evaluated with the DHARMA package (Hartig 2022).

Validating future predictive ability

As a first validation, we leveraged the natural recursive element of our data to quantify the ability of our models to predict the future distribution of ages. For each of the species - age models constructed above, we made predictions to the survey locations in the following year (e.g. age 3 hake in years 2007 – 2018 used to predict the distribution of age 4 hake in 2008 – 2019). We related the predicted probabilities of occurrence $\hat{p}_{s,t}$ to observations by multiplying predictions by the total number of fish sampled for ageing, $\hat{n}_{s,t} = N_{s,t} \cdot \hat{p}_{s,t}$. We

quantified the relationship between predictions and observations by fitting a simple Poisson GLM, where counts of fish of age $a+1$ in year $t+1$ were treated as the response and $\log(\lambda_{s,t}) = \beta_0 + \beta_1 \cdot \log(\hat{n}_{s,t})$. The exponentiated slope parameter $\exp(\beta_1)$ represents a the change in expected counts that would be expected from a 1-unit change in $\log(\hat{n}_{s,t})$, and can roughly be interpreted as an estimate of relative survival.

As a second validation, we compared aggregated predictions across all ages to a SDM model that used total biomass as a response.

Links to fisheries data?

Results

One useful output from our spatiotemporal models is the estimated spatial field, ω_s , representing the average spatial distribution across years. Results from our model of North Pacific hake show a concentration of age-1 hake near the coast, and more of latitudinal north - south break for 2 - 3 year old hake (with higher occurrences in the South) Figure 1. As expected, age 4 hake appear to have a northern distribution, though the gradient is less clear than for ages 2 - 3. For sablefish, our models estimate a concentration of age 0 - 2 sablefish in shallow water, with a deeper distribution by age 3 (Figure 2). Latitudinally, there are several areas in central Oregon that appear to have consistently lower occurrences than average for older age classes (Figure 2).

Estimates of age composition by year may be useful in tracking the distribution of cohorts through time, as well as examining variability in distributions of particular age classes over time. For example, results from our models can be used to project the spatial distributions of strong hake cohorts (2010, 2014, Figure 3) and strong sablefish cohorts (2013, 2016, 2020, Figure 4). [expand on this]

When comparing the spatial and spatiotemporal parameters across ages and species, we find that for most ages, we see a decrease in the spatial and spatiotemporal variances as well as a decrease in the spatial range (Figure 8). Spatial and spatiotemporal variances control the magnitude of the peaks and valleys in estimated latent fields, suggesting that the fields estimated for older fish are less variable. Similarly, the spatial range determines the distance at which locations are functionally independent (for the Matern, $\rho \sim 0.13$). The decline in range with age implies that spatial distributions also become more concentrated as a function of age (Figure 8).

Our 1-step ahead validation analyses indicate that there is a stronger association between older ages compared to younger ages, and this effect is strongest for sablefish (Figure 5).

Figure Figure 6 aggregates occurrence probabilities for age 0 - 2 sablefish, potentially useful to identify hotspots

Similarly Figure Figure 7 shows occurrence probabilities for age 5 sablefish

Discussion

Limitations

- not predicting abundance (N) – just proportions
- We constructed single-age models, and this could be done in a multivariate (multinomial) framework – but our approach allows for greater flexibility in letting relationships change by age. Tradeoff is that it doesn't impose sum to 1 constraint, binomials are independent

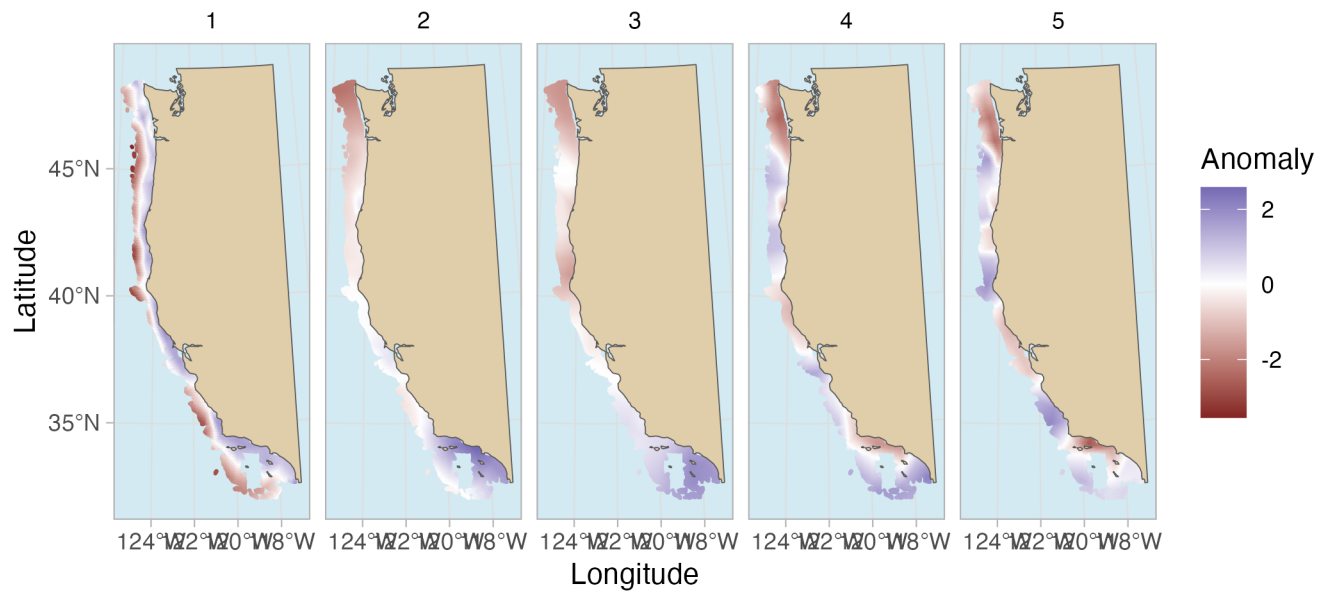


Figure 1: Estimated spatial anomalies (shared across all years) for Pacific hake ages 1 – 4.

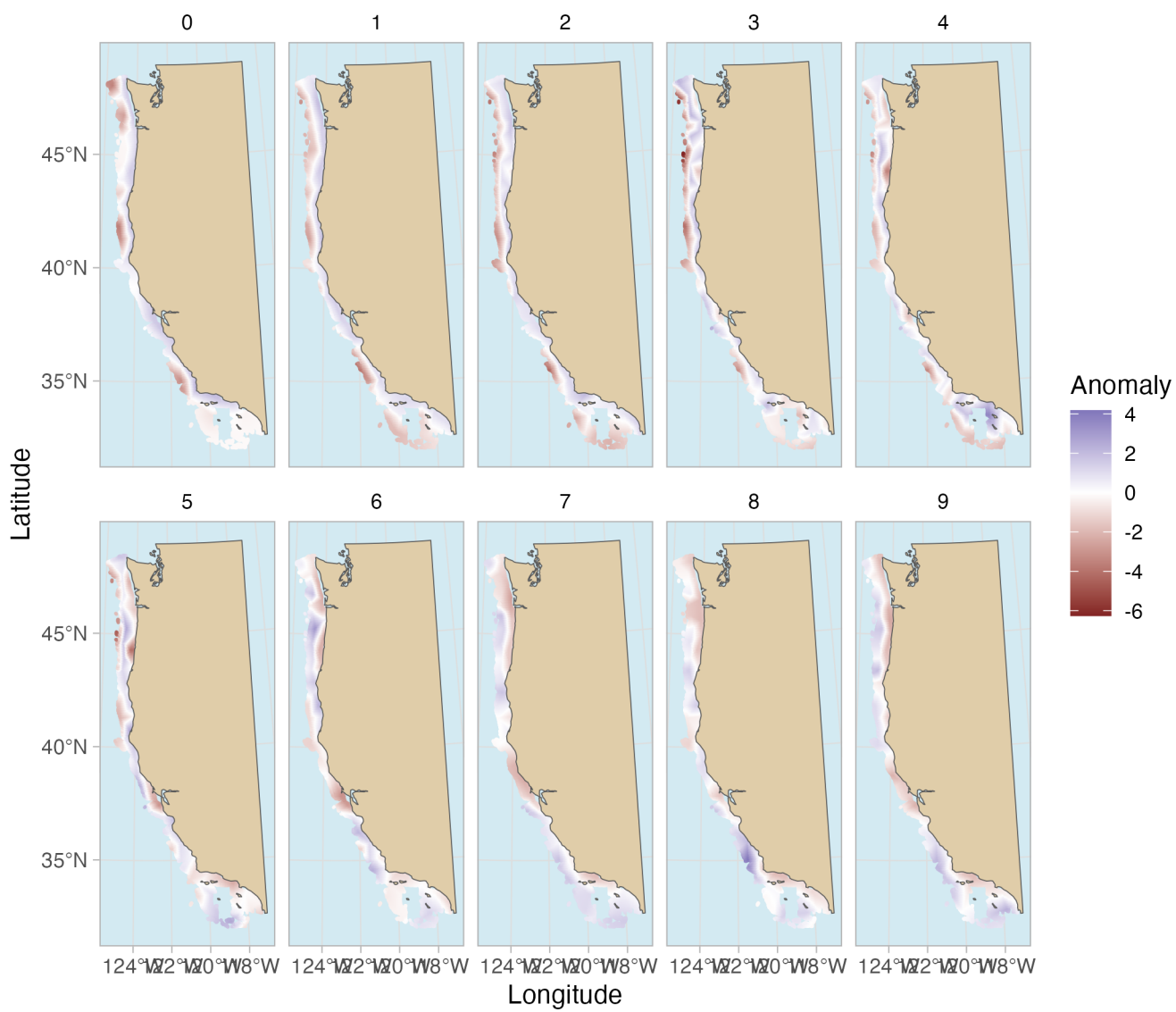


Figure 2: Estimated spatial anomalies (shared across all years) for sablefish ages 0 – 5.

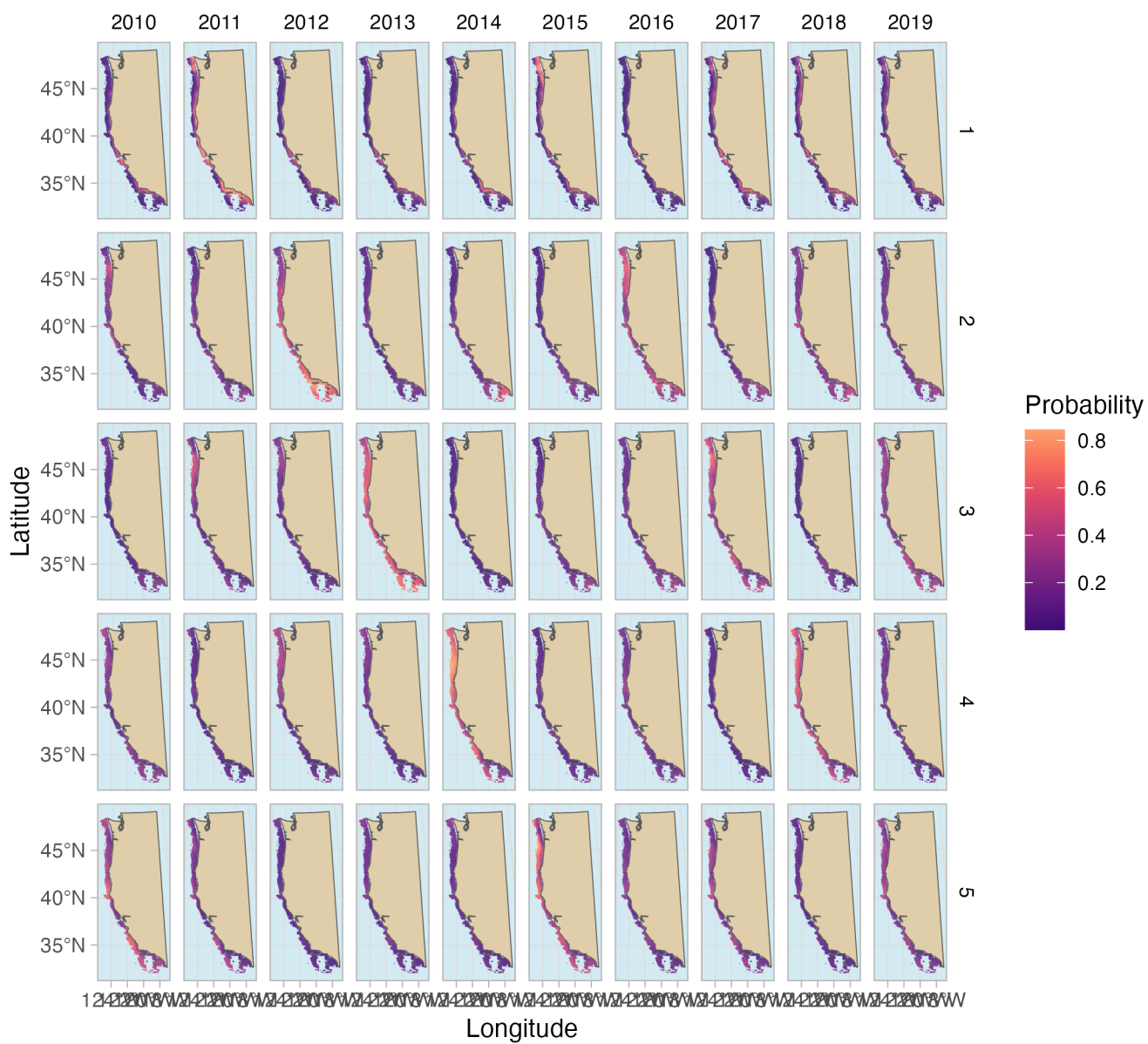


Figure 3: Estimated spatial composition for Pacific hake; for plotting purposes only ages 1 - 5 and the last 10 years of available predictions are shown.

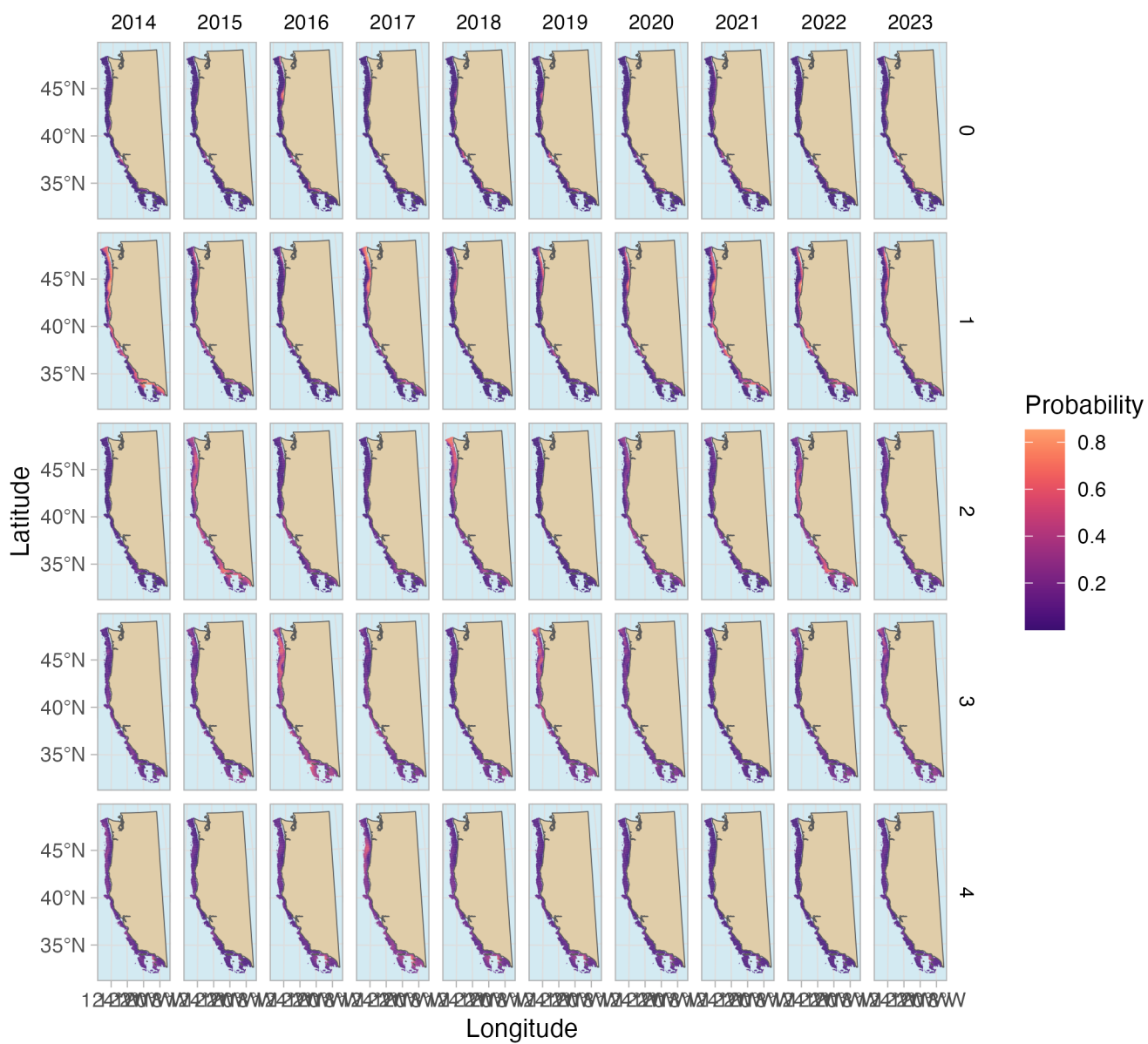


Figure 4: Estimated spatial composition for sablefish; for plotting purposes only ages 0 - 4 and the last 10 years of available predictions are shown.

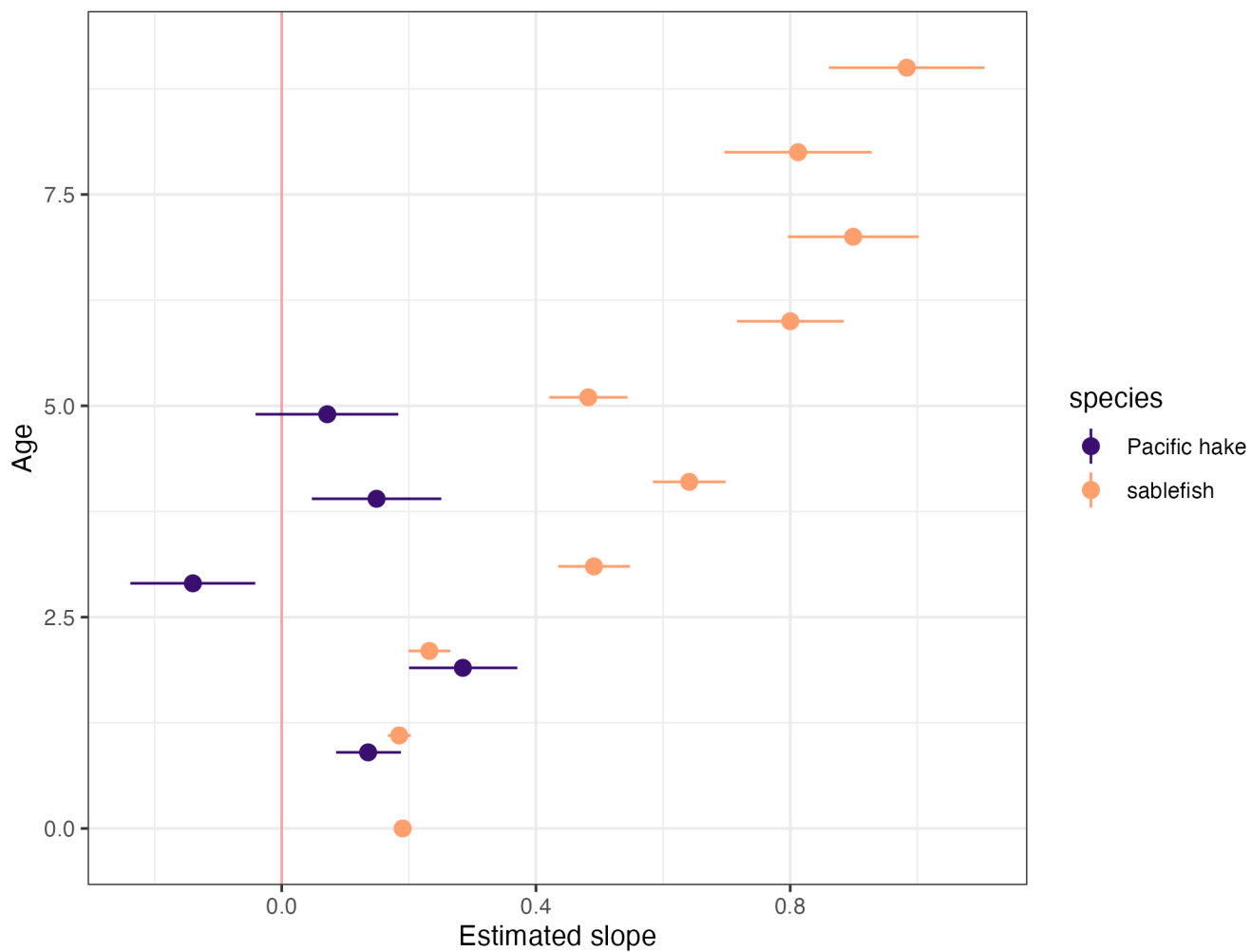


Figure 5: Estimated coefficients relating predicted densities of age a fish in year t to observed numbers the following year. The coefficients are interpreted with a multiplicative effect, such that a coefficient of 0.5 translates to an effect of $\exp(0.5) = 65\%$ increase on the expected count in the future.

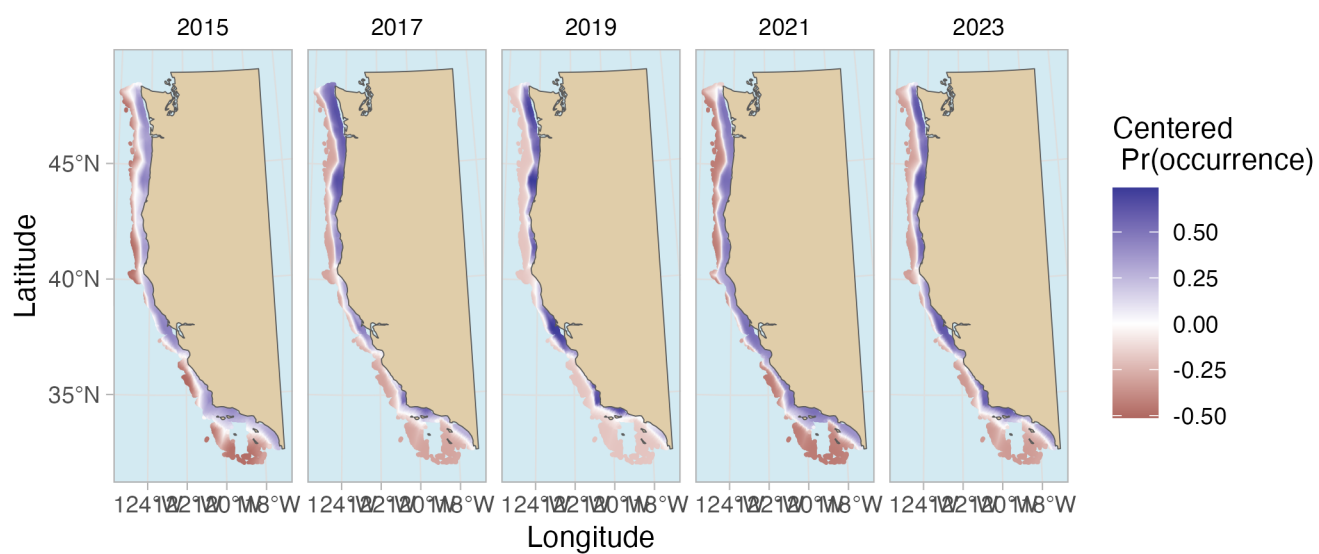


Figure 6: Estimated probability of occurrence for age 0 – 2 sablefish.

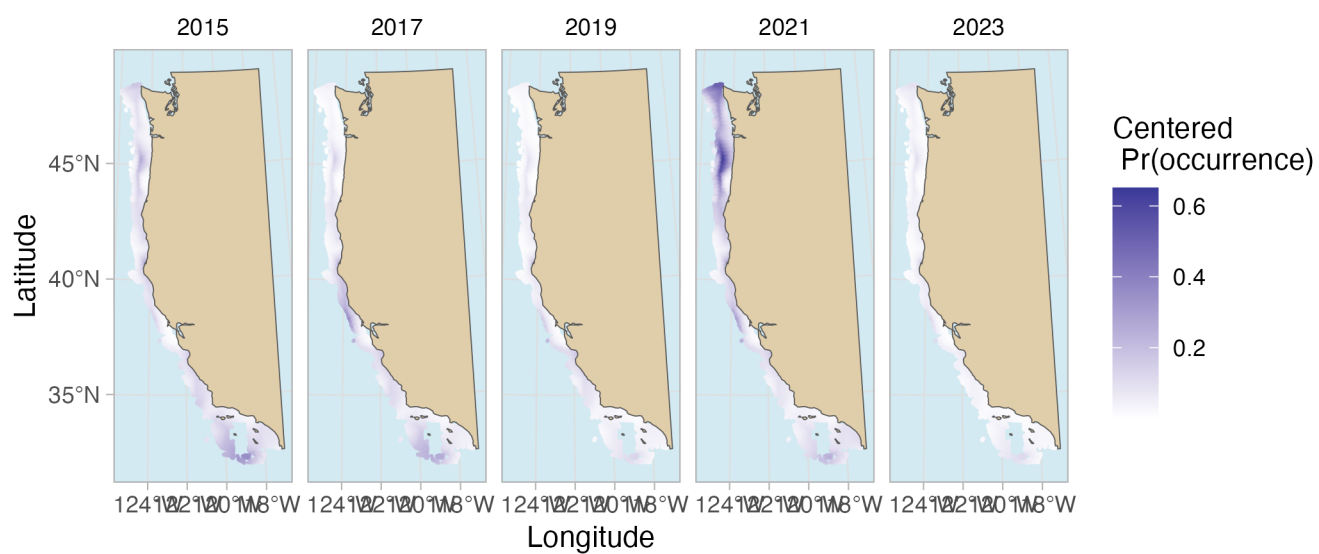


Figure 7: Estimated probability of occurrence for age 5 sablefish.

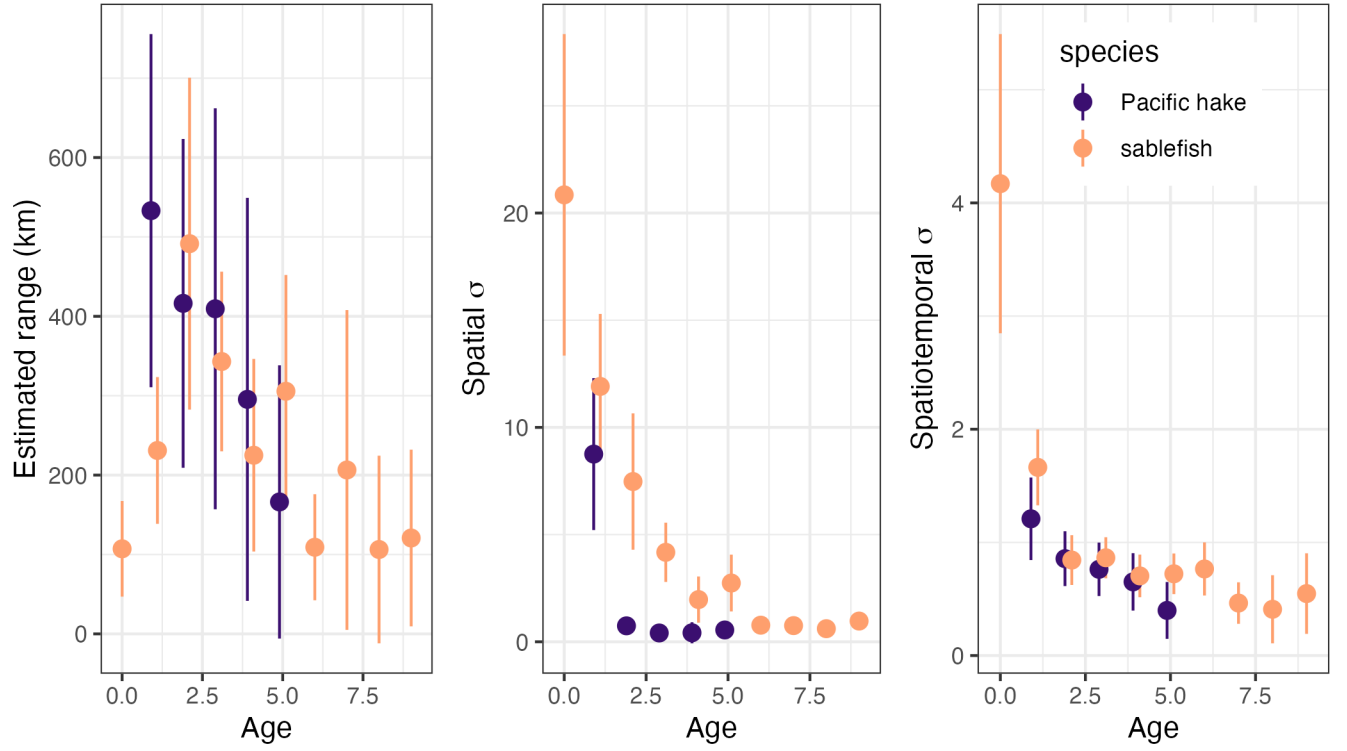


Figure 8: Estimated spatial parameters (range, and spatial and spatiotemporal standard deviations). Points represent mean estimates; lines represent 95% confidence intervals

References

- Anderson, Sean C., Eric J. Ward, Philina A. English, Lewis A. K. Barnett, and James T. Thorson. 2024. “sdmTMB: An r Package for Fast, Flexible, and User-Friendly Generalized Linear Mixed Effects Models with Spatial and Spatiotemporal Random Fields.” *bioRxiv*. <https://doi.org/10.1101/2022.03.24.485545>.
- Cao, Jie, James T. Thorson, R. Anne Richards, and Yong Chen. 2017. “Spatiotemporal Index Standardization Improves the Stock Assessment of Northern Shrimp in the Gulf of Maine.” *Canadian Journal of Fisheries and Aquatic Sciences* 74 (11): 1781–93. <https://doi.org/10.1139/cjfas-2016-0137>.
- Ciannelli, Lorenzo, Kung-Sik Chan, Kevin M. Bailey, and Nils Chr. Stenseth. 2004. “NON-ADDITIVE EFFECTS OF THE ENVIRONMENT ON THE SURVIVAL OF A LARGE MARINE FISH POPULATION.” *Ecology* 85 (12): 3418–27. <https://doi.org/10.1890/03-0755>.
- Elith, Jane, and John R. Leathwick. 2009. “Species Distribution Models: Ecological Explanation and Prediction Across Space and Time.” *Annual Review of Ecology, Evolution, and Systematics* 40 (1): 677–97. <https://doi.org/10.1146/annurev.ecolsys.110308.120159>.
- Fredston, Alexa, Malin Pinsky, Rebecca L. Selden, Cody Szuwalski, James T. Thorson, Steven D. Gaines, and Benjamin S. Halpern. 2021. “Range Edges of North American Marine Species Are Tracking Temperature over Decades.” *Global Change Biology* 27 (13): 3145–56. <https://doi.org/10.1111/gcb.15614>.
- Guisan, Antoine, and Wilfried Thuiller. 2005. “Predicting Species Distribution: Offering More Than Simple Habitat Models.” *Ecology Letters* 8 (9): 993–1009. <https://doi.org/10.1111/j.1461-0248.2005.00792.x>.
- Hartig, Florian. 2022. *DHARMA: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models*. <https://CRAN.R-project.org/package=DHARMA>.
- Hilborn, Ray, and Carl J. Walters. 1992. *Quantitative Fisheries Stock Assessment*. Springer US. <https://doi.org/10.1007/978-1-4615-3598-0>.
- Johnson, Kelli F., James T. Thorson, and André E. Punt. 2019. “Investigating the Value of Including Depth During Spatiotemporal Index Standardization.” *Fisheries Research* 216 (August): 126–37. <https://doi.org/10.1016/j.fishres.2019.04.004>.
- Kristensen, Kasper, Anders Nielsen, Casper W. Berg, Hans Skaug, and Bradley M. Bell. 2016. “TMB: Automatic Differentiation and Laplace Approximation.” *Journal of Statistical Software* 70 (5). <https://doi.org/10.18637/jss.v070.i05>.
- Laman, Edward A., Christopher N. Rooper, Kali Turner, Sean Rooney, Dan W. Cooper, and Mark Zimmermann. 2018. “Using Species Distribution Models to Describe Essential Fish Habitat in Alaska.” *Canadian Journal of Fisheries and Aquatic Sciences* 75 (8): 1230–55. <https://doi.org/10.1139/cjfas-2017-0181>.
- Lewison, Rebecca L., Sloan A. Freeman, and Larry B. Crowder. 2004. “Quantifying the Effects of Fisheries on Threatened Species: The Impact of Pelagic Longlines on Loggerhead and Leatherback Sea Turtles.” *Ecology Letters* 7 (3): 221–31. <https://doi.org/10.1111/j.1461-0248.2004.00573.x>.
- Lindgren, Finn, and Håvard Rue. 2015. “Bayesian Spatial Modelling with *R-INLA*.” *Journal*

- of *Statistical Software* 63 (19). <https://doi.org/10.18637/jss.v063.i19>.
- Lindgren, Finn, Håvard Rue, and Johan Lindström. 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): 423–98. <https://doi.org/10.1111/j.1467-9868.2011.00777.x>.
- Melo-Merino, Sara M., Héctor Reyes-Bonilla, and Andrés Lira-Noriega. 2020. “Ecological Niche Models and Species Distribution Models in Marine Environments: A Literature Review and Spatial Analysis of Evidence.” *Ecological Modelling* 415 (January): 108837. <https://doi.org/10.1016/j.ecolmodel.2019.108837>.
- Methot, Richard D., and Chantell R. Wetzel. 2013. “Stock Synthesis: A Biological and Statistical Framework for Fish Stock Assessment and Fishery Management.” *Fisheries Research* 142 (May): 86–99. <https://doi.org/10.1016/j.fishres.2012.10.012>.
- Phillips, Natasha D., Neil Reid, Tierney Thys, Chris Harrod, Nicholas L. Payne, Cheryl A. Morgan, Hannah J. White, Siobhán Porter, and Jonathan D. R. Houghton. 2017. “Applying Species Distribution Modelling to a Data Poor, Pelagic Fish Complex: The Ocean Sunfishes.” *Journal of Biogeography* 44 (10): 2176–87. <https://doi.org/10.1111/jbi.13033>.
- R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Rijnsdorp, Adriaan D., Stefan G. Bolam, Clement Garcia, Jan Geert Hiddink, Niels T. Hintzen, P. Daniel van Denderen, and Tobias van Kooten. 2018. “Estimating Sensitivity of Seabed Habitats to Disturbance by Bottom Trawling Based on the Longevity of Benthic Fauna.” *Ecological Applications* 28 (5): 1302–12. <https://doi.org/10.1002/eap.1731>.
- Thorson, James T., Andrew O. Shelton, Eric J. Ward, and Hans J. Skaug. 2015. “Geostatistical Delta-Generalized Linear Mixed Models Improve Precision for Estimated Abundance Indices for West Coast Groundfishes.” *ICES Journal of Marine Science* 72 (5): 1297–1310. <https://doi.org/10.1093/icesjms/fsu243>.
- Werner, E E, and J F Gilliam. 1984. “The Ontogenetic Niche and Species Interactions in Size-Structured Populations.” *Annual Review of Ecology and Systematics* 15 (1): 393–425. <https://doi.org/10.1146/annurev.es.15.110184.002141>.