

Incorporating spatial distribution at age improves predictive ability of SDMs

Abstract

Introduction

Why hake and sablefish? 2 species with strong gradients in distribution at age

Methods

Data

On the west coast of the USA, the West Coast Groundfish Bottom Trawl Survey (WCGBTS) has been an annual survey conducted from 2003 - present. The WCGBTS 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, effort, gear, seasonal, and spatial coverage have remained relatively constant within the period we analyse (data from 2019 also exists but is not included because of a planned halving of survey effort compared to previous years). Though the survey samples hundreds of species, we focused our modelling on the 36 species sampled with the highest densities.

We concentrated our analysis on two of the well sampled species from the WCGBTS, North Pacific hake and sablefish. [more on biological gradients] While otoliths are sampled continuously during the WCGBTS 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 – 6 (56% and 63% of the total distribution of ages, respectively). For each species - age combination, we constructed a unique spatiotemporal model fit to all years except the last available year for each species. We first aggregated data across individual fish at the haul level, summarizing 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 (with 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 $\epsilon_t \sim \text{MVNormal}(0, \Sigma_\epsilon)$

Estimation was performed via sdmTMB (Anderson et al) and using TMB / INLA SPDE etc...

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 and intensity of age 4 hake in 2008 – 2019). We related the predictions $\hat{p}_{s,t}$ to observations by multiplying predictions by the total number of fish sampled for ageing, $\hat{y}_{s,t} = N_{s,t} \cdot \hat{p}_{s,t}$. We quantified the relationship between predictions and observations by fitting a simple Poisson GLM, with counts of fish of age $a + 1$ in year $t + 1$ treated as the response and log predictions $\log(\hat{y}_{s,t})$ as the covariate. The slope of this GLM represents a 1:1 change in the response, or the percent change in predicted fish of age a in year t that would be needed to achieve a similar percent change in the response.

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

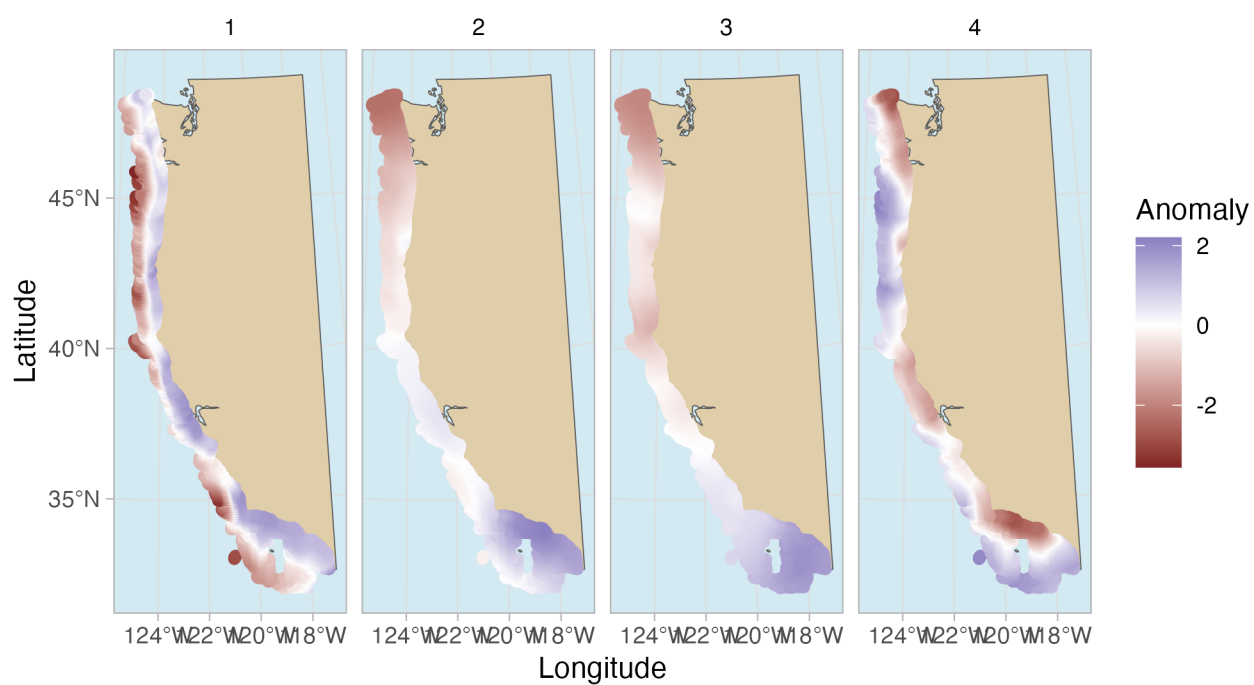


Figure 1: Estimated spatial anomalies (shared across all years) for Pacific hake ages 1 – 4.

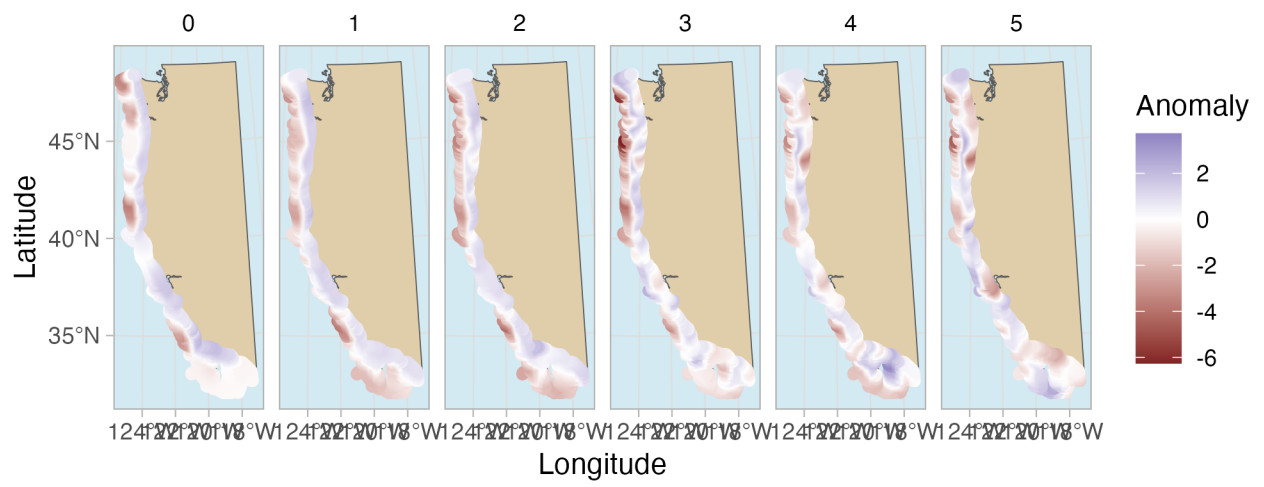


Figure 2: Estimated spatial anomalies (shared across all years) for sablefish ages 0 – 5.

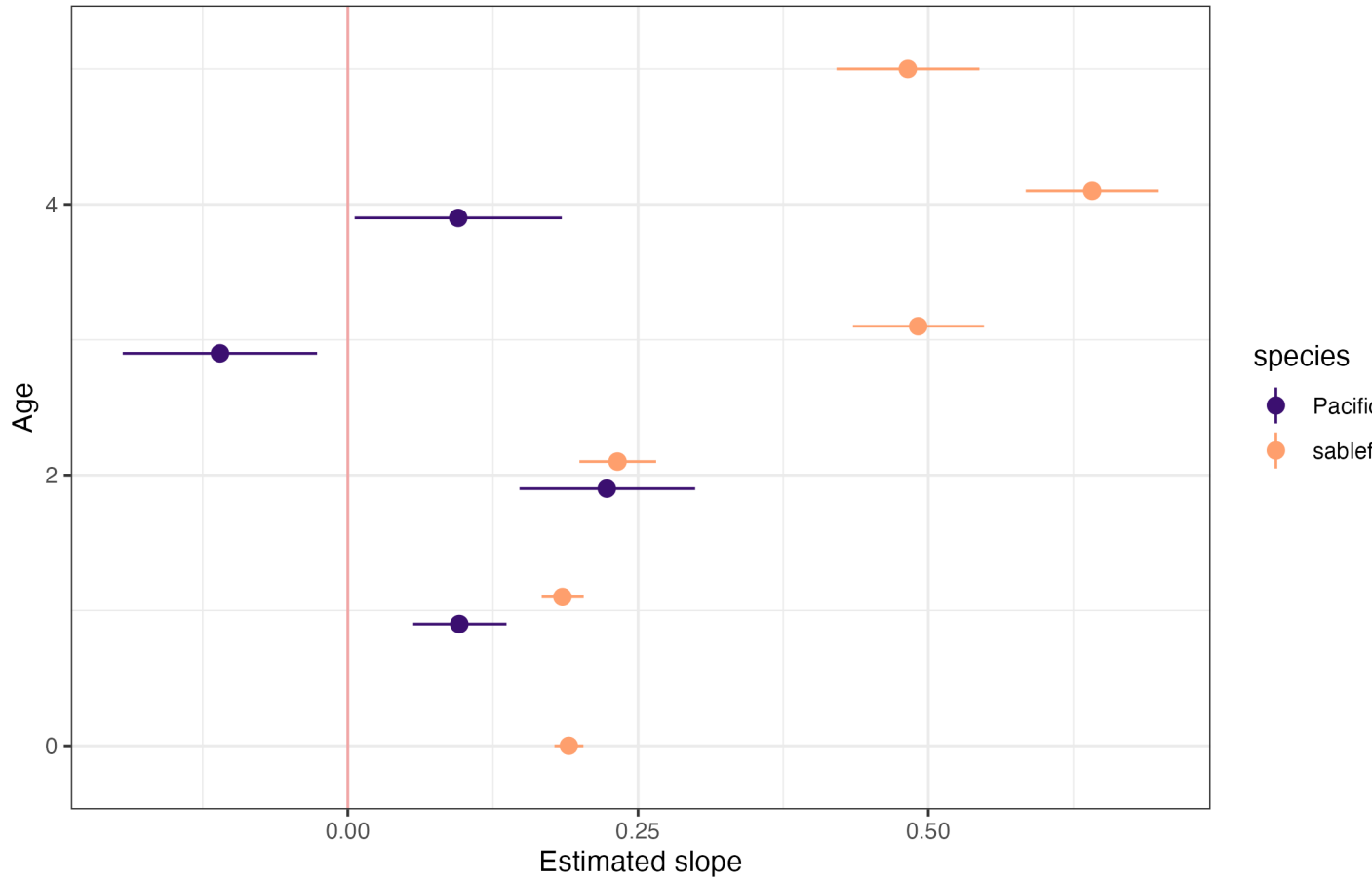


Figure 3: Estimated coefficients relating predicted densities of age a fish in year t to observed numbers the following year. The coefficients are interpreted with a multiplicative effect, such that a coefficient of 0.5 translates to an effect of $\exp(0.5) = 65\%$ increase on the expected count in the future.