Estimating shifts in squid distribution

*Data*

We compiled data from

*Spatiotemporal model*

To estimate shifts in the distribution of squid between 1998 and 2019 in response to biological (e.g., chlorophyll concentration (xx/mm), Chl a) and climate forces (e.g., temperature), we analyzed catch-per-unit-effort (CPUE) data from fisheries independent surveys conducted by the National Oceanic and Atmospheric Association (NOAA) between 1998 and 2019. We used a delta-generalized linear model (delta-glmm) to estimate the probability of encountering squid during a survey, and the probability of positive catches if squid were encountered

Equation 1.

Where for the ith sample, is observed number of squid captured, is the probability of positive catches, is the expected number of squid captured given positive catches, is the effort offset (i.e., distance fished times the average opening of the net (0.05 km)), is the observed error not explained by biological or environmental covariates, or random variation in the spatiotemporal distribution of the catches, and g is a probability distribution describing the positive catches (e.g., gamma, log-normal, etc.).

We used a variance-autoregressive spatiotemporal (VAST) model implemented in the R package VAST to partition the sources of variability in the catches based on the fixed effects associated with the biological and environmental covariates, and random spatial and spatiotemporal variation for the encounter probabilities and positive catch rates. Thorson (2019) provides a detailed description of the statistical properties of VAST models, as well as the numerous decisions analyst implementing VAST models must make concerning the type and quality of the data they are using. Table 2 provides the fifteen decisions for our squid analysis as they relate to the decision tree provide by Thorson (2019).

Briefly, and for the purposes of our data, the VAST models include linear estimators for the encounter probability (Equation 2) and positive catches (Equation 3) based on the i) intercepts, ii) covariates, and iii) spatial and spatiotemporal processes,

Equation 2.

Equation 3.

Based on an initial analysis of the size distribution data (Supplemental figure xx), we categorized squid lengths in two bins, mm and 80 mm. The intercept parameters represent estimates of the annual changes in the encounter probability and positive catches, respectively, for category in year for the ith observation.

The spatial and spatiotemporal random effects describe the residual variance not explained by the fixed intercepts or covariates for location of the ith observation related to factor f. The VAST package allows us to estimate the correlation between the spatial fields for each length bin for both the spatial and spatiotemporal processes using dynamic factor analysis. Grouping the length bin categories by different factors, where the number of factors is less than or equal to the number of categories (), is done with the loadings matrices and a method referred to as dynamic factor analysis. A complete description of dynamic factor analysis is available in Holmes et al. and Thorson et al. . Our current analysis includes only a single species (i.e., squid) and two categorical length bins. Therefore the number of factors () in the loadings matrices to address how these two length bins are correlated in space is simply equal to two . However, future analyses may include questions concerning how multiple species (competitors or predators) and additional length bins are correlated in space, and the statistical framework of the VAST package allows for easy integration of additional data to address these questions.

The parameters describe the change in the catchability of squid based on the kth covariate related to the ith observation . Because the covariates are related to the observation, there is no subscript for the encounter or positive catch models. Catchability covariates are different from density covariates. Density covariates are observed at every location in every year and are useful for extrapolation and forecasting. The catchability covariates are anything that could affect catch rates but not density. In our case, we have data on the 3m depth measurements of temperature, chlorophyll a concentrations, and salinity which could all plausibly affect the vertical distribution of squid in search for suitable habitat and prey concentrations.

*Model estimation, validation and selection*

Using libraries in the TMB (Kristensen et al. 2015) package in R, the parameters of VAST model are estimated by maximizing the marginal likelihood with respect to the fixed effect, while integrating out the random effect using a Laplace approximation. We use both AIC (Akaike 1974) and k-fold cross-validation to explore which single model best explains the data.

Results

Tables

Table 1. Description of model data, parameters, variables, and subscripts.

|  |  |
| --- | --- |
| Subscripts | Description |
|  | ith observation |
|  | References observations with positive catches |
|  | References observations with zero catch |
| Indexes |  |
|  | Year |
|  | Length category |
|  | Station were catches occurred |
|  | Number of years |
|  | Number of observations |
|  | Number of categories |
|  |  |
| Fixed effects |  |
|  | Intercepts for zero () and positive catches (), for the category () and year () of the ith observation |
| Random effects |  |
|  |  |
|  |  |

Table 2. VAST decision tree. The object data is based on the xx.csv.

|  |  |  |
| --- | --- | --- |
| Decision | Description | VAST |
| 1) Spatial domain used when calculating derived quantities | We use the west coast of the US from San Francisco Bay to the northern tip of Washington State. We divide the coast into four strata, with ‘coastwide’ as an aggregate for the three states | strata.limits <- data.frame(  'STRATA' = c("Coastwide","CA","OR","WA"),  'north\_border' = c(49.0, 42.0, 46.0,49.0),  'south\_border' = c(32.0, 32.0, 42.0, 46.0)  )) |
| 2) Which categories (species/sizes) to include | We included two size categories, and | c\_i <- data$c\_i |
| 3) Identify whether to analyze encounter, abundance, and/or biomass-sampling data | The data are the number of squid captured. | b\_i <- data$b\_i |
| 4) Including spatial and/or spatiotemporal variation | The full model has spatial and spatiotemporal process for the encounter probability and positive catches | FieldConfig = c("Omega1" = 1, "Epsilon1" = 1, "Omega2" = 1, "Epsilon2" = 1) |
| 5) Choosing the spatial smoother and resolution | We used a “mesh” grid, with 200 nodes, and assume geometric anisotropy | Mesh.Method <- "Mesh"  n\_x <- 200  Aniso <- TRUE |
| 6) Choosing the number of spatial and spatio-temporal factors | We evaluate a full rank model where each length category has its own covariance matrix | FieldConfig <-c(Omega1 =2, Epsilon1 =2, Omega2 =2, Epsilon2 =2) |
| 7) Specifying temporal correlation on model components | We do not assumed any temporal correlation in the intercepts or spatiotemporal processes | RhoConfig= c("Beta1"=0, "Beta2"=0, "Epsilon1"=0, "Epsilon2"=0) |
| 8) Including density covariates as a semi-parametric model | We have no density dependent covariates for the model |  |
| 9) Accounting for catchability covariates and confounding variables | We include catchability covariates for temperature, chlorophyll a, and salinity at 3m | Q\_ik <- raw[,c('x3m\_Temp','x3m\_Salinity','x3m\_Chl')] |
| 10) Treating area swept as a catchability covariate or offset | Distance towed | a\_i <- raw$effort |
| 11) Including vessel effects as overdispersion | We do not include any vessel effects | OverdispersionConfig <- c("Eta1"=0, "Eta2"=0) |
| 12) Choosing among link functions and distributions | We assume a delta-GLMM with a log-normal distribution for the positive catches | ObsModel <- c(2,0) |
| 13) Derived quantities | Center of gravity  Annual estimate of squid density | Options = c(SD\_site\_density = 0  ,SD\_site\_logdensity = 0  ,Calculate\_Range = 1  ,Calculate\_evenness = 0  ,Calculate\_effective\_area = 1  ,Calculate\_Cov\_SE = 0  ,Calculate\_Synchrony = 0  ,Calculate\_Coherence = 0) |
| 14) Bias correction for derived quantities |  |  |
| 15) Model selection | AIC |  |