Estimating shifts in squid distribution

**Introduction**

Our goal was to make inference about the abundance and distribution of market squid along the west coast of the United States using a geostatistical model that accounts for biases in the survey design. Specifically, our model estimated the abundance, the area occupied, and the center of gravity (COG; Thorson et al. 2016) of the distribution for squid in the upper xx meters of the water column.

**Methods**

*Data*

We obtained squid catch data from fishery independent surveys conducted by the Northwest Fisheries Science Center (NWFSC) from 1998 to 2019, and from 2010 to 2016 for the Southwest Fisheries Science Center (SWFSC) (Figure 1). All trawls were conducted during the day. The locations of the trawls for the NWFSC and SWFSC were a combination of predetermined transects (citation, Emmett or Peterson) and *ad hoc* adjustments in years with low catches or limited research funds.

Trawl effort was defined by 30 minutes tows; however, for our purposes we used area swept as calculated by the GPS coordinates between start and end points and the effective width of the trawl opening – estimated to be approximately 28 meters (Brodeur et al., 2005; Emmett et al., 2006; Harding et al., 2011). The total number of squid in each trawl were counted when reasonable, or the total catch was extrapolated from a sample weight for a known volume of squid when catches exceeded the science crew’s capacity to finish processing fish between trawls. Up to 50 individual squid lengths were recorded for each trawl: we assume the length frequency of the entire catch was the same as the length frequency of the sample. Additionally, biophysical data such as the temperature (), chlorophyll concentration (mg m-3 ???), and salinity (ppm) were collected at three meters before each trawl.

Potential biases exist based on the gear, vessel, and crew that were used. For the NWFSC, the F/V Ocean Star conducted surveys from 1998 to 2000, and the F/V Frosti conducted surveys from 2001 to 2019. During that time, any changes in the science and vessel crew leads occurred after overlap with previous crews. The F/V XX was used to conduct the SWFSC surveys from 2010 to 2016. Similar to the NWFSC surveys, there was significant overlap between years for both the science and vessel crews. The net was a 264 Nordic rope trawl (see NMFS (2008) and Krutzikowsky and Emmett (2005) for a complete description). To prevent the capture of non-target species, a marine mammal excluder device was added to the SWFSC surveys from 2012 to 2016, and NWFSC surveys from 2014 to 2019. During the initial year of deployment the MMED was placed in an upward position, but was changed to a downward position in subsequent years. Paired trawls, with and without the MMED were conducted to evaluate the effects of the MMED. Squid catches declined by 12% and 52% when the MMED was in upward and downward position, respectively (Wainwright et al., 2019). Based on our best available knowledge when addressing these potential biases, we chose to assume that there is no crew or vessel bias between years. One option would be to compare the spatiotemporal synchrony for the NWFSC and SWFSC surveys; however, with no spatial overlap, and only a small number of year of sampling for the SWFSC, models comparing the temporal synchrony of the two surveys failed to converge during preliminary analyses. Therefore, we treated the catches from the SWFSC and NWFSC as a single survey. Future work may consider methods for disaggregating these two surveys to examine biases, but that is currently beyond the scope of this analysis. Finally, we adjusted the total squid catches *a prior* for years with MMED using the catch ratios for the upward and downward position (12% and 52%, respectively) found in Wainwright et al. (2019).

*Spatiotemporal model*

To account for the unbalanced design of the two surveys, we used a geostatistical model with random processes describing the effects of space and time on the estimated catches of market squid. The delta-generalized linear model (delta-glmm) consists of two parts, the probability of encountering squid during a survey, and the probability of positive catches if squid were encountered

Equation 1.

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, 0.085 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.). The joint probability for the observations includes a probability for the encounters and positive catches. Because we did not have true counts of the squid (i.e., the number of squid were often extrapolations based on weights or volumes) we chose the gamma distribution (i.e., , where is equal to the coefficient of variation for the observations) rather than the Poisson or negative-binomial distribution. For the encounter probability, we assumed a Bernoulli distribution ().

We used a geospatial model that included 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 found little evidence of distinct modes in the data that would support a dynamic factor analysis. Market squid grow quickly, spawn continuously, and only live one year (citation) making it unlikely that an annual survey would detect cohort differences. The vector intercept parameters, and , are the annual changes in the encounter probability and positive catches, respectively, for the ith observation, and they maybe estimated as fixed effects or random effect based with auto-regressive process (AR1). For example, using the intercepts for the encounter probability, the AR1 process is defined as , where is the correlation between time steps, and is the variance, with a similar process but different subscripts used to explain the positive catches.

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. The catchability covariates are anything that could affect catch rates but not density. In our case, we have data on the temperature, chlorophyll a concentrations, and salinity at a depth of 3m; all these covariates could affect the vertical distribution and catch rates of squid as they search for suitable habitat and prey concentrations. Catchability covariates are different from density covariates: density covariates are observed at every location in every year and are useful for extrapolation and forecasting. Although we are not concerned with forecasting, future research examining how environmental forces affect the distribution and abundance of squid could be incorporated into our R code and model structure.

The spatial and spatiotemporal random effects describe the residual variance not explained by the fixed intercepts or covariates for location of the ith observation. The probability distributions for the spatial and spatiotemporal effects were assumed to be multivariate normal for the both encounter probability and the positive catches, where the covariance between locations is governed by a Matern function with geometric anisotropy (see Thorson 2019 for a detailed description). The anisotropy allows the decorrelation along the north-south and east-west to differ. Estimating the anisotropy adds two additional parameters to the model, but this feature is relevant since squid are known to be associated shelf features. The spatiotemporal variability may also include auto-correlation, such that and . Where and are vectors of the spatial variability for the encounter rate and positive catches during year t, and are the respective correlations for the AR1 processes, and and are the covariance matrices which include a Matern function describing the decorrelation across space (See Thorson (2019) for a complete description).

*Model estimation, validation and selection*

To estimate the parameters of the model and partition the variance associated with different random effects, we used the variance-autoregressive spatiotemporal (VAST) package in the R (Thorson 2016). The VAST package uses the Template Model Builder libraries (TMB package; Kristensen et al. 2015) for R to maximize the marginal likelihood of the fixed effects, while integrating out the random effects (geospatial and temporal variation) using a Laplace approximation.

We followed Thorson’s fifteen step decision tree when implementing the spatiotemporal model in the VAST package when exploring various model structures. Given the large number of random effects needed to estimate the geospatial effects in the model, the VAST package uses the matrix of second derivatives for the marginal likelihood to check that the fixed-effects are identifiable and the parameters of the model are estimable. Where comparing among models that included or did not include the catchability coefficients, and or spatial and spatiotemporal processes, we used AIC (Akaike 1974). Finally, for the best model chosen based on AIC, we k-fold cross-validation with 10 folds to determine how well the model will perform on an unknown dataset.

**Results**

We were able to fit geostatistical models to fisheries independent squid catches between 1998 and 2019, and the top model with the lowest AIC included spatial and spatiotemporal variation in the for the encounter probability and density of squid, suggesting a large amount of spatiotemporal variability between years for the two observation processes (Table xx). Hotspots of squid density are observed Additionally, among the top two models which both include spatial and spatiotemporal variation, the model with the catchability coefficients was significantly better than model without ().

Results from the geostatistical suggest that the range of the center of gravity (COG) of the survey catches has moved between 150 km north and south and 30 km east and west (Figure 4); but there has been little trend in those shifts over time. Furthermore, the large shift in the distribution during 2015 was due to increased catches across the entire Oregon and Washington coast, with a particularly large hotspot of catches around the Oregon/California border (Figure 3 and Figure 4).

When we compared the coastwide catches to disaggregated catches for California, Oregon, Washington, and the coastwide (Figure 5), we found strong synchrony between each area, providing further evidence that squid populations are increasing their traditional and marginal habitat. With the largest increases coming xx and the lowest coming in xx. We did find an increase in the range for both size classes of squid: between 1998 and 2019, small and large squid expanded their ranges by xx and xx percent, respectively, in the surveyed water (Figure 6).

Tables

Table 1. Description of model data, parameters, variables, and subscripts. Not listed are the fixed effect parameters governing the spatial and spatiotemporal random processes, or the computed quantities used to estimate the anisotropy matrix. See Thorson (2019) for the complete description of these VAST equations.

|  |  |  |
| --- | --- | --- |
| Subscripts |  | Description |
|  |  | The ith observation associated with a particular survey tow |
|  |  | Positive catches |
|  |  | Zero catches |
| Indexes |  |  |
|  |  | Year |
|  |  | Station were catches occurred |
|  | k | Environmental covariate (e.g., salinity, chlorophyll, temperature) |
| Fixed effects |  |  |
|  |  | Intercepts for zero () and positive catches (), for year () of the ith observation |
|  | and | Coefficient relating covariate k to the presence (p) and density (r) of squid catches |
|  | and | Temporal autocorrelation for the intercepts |
|  | and | Temporal correlation for the spatiotemporal processes describing the encounter rate and positive catches |
|  |  | A measure of precision for gamma distribution observations where is equal to the coefficient of the variation |
|  | and | Variability of the AR1 processes describing the intercept parameters |
| Random effects |  |  |
|  |  | Spatial variability for the presence (p) and density (r) of squid catches |
|  |  | Spatiotemporal variability for the presence (p) and density (r) of squid catches |
| Covariates |  |  |
|  |  | The kth environmental covariate observed during the ith survey tow |

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 | 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 | a single size class, or two size classes and , or | c\_i <- data$c\_i |
| 3) Identify whether to analyze encounter, abundance, and/or biomass-sampling data | The extrapolated number of squid captured. | b\_i <- data$b\_i |
| 4) Including spatial and/or spatiotemporal variation | The full model has spatial and spatiotemporal process for the encounter probability and positive catches for each size category. See step 6, for multivariate model. | FieldConfig <- c("Omega1" = 2, "Epsilon1" = 2, "Omega2" = 2, "Epsilon2" = 2) |
| 5) Choosing the spatial smoother and resolution | We used a “mesh” grid, with 200 nodes, and assume geometric anisotropy – east/west and north/south deviates are not symmetric. | 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. The “2” represents separate spatiotemporal process for the two size categories. | FieldConfig <-c(Omega1 =2, Epsilon1 =2, Omega2 =2, Epsilon2 =2) |
| 7) Specifying temporal correlation on model components | We test whether there is “1” or is not “0” temporal correlations in the intercepts or spatiotemporal processes. Default is not temporal correlation. | 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 | Area towed, distance towed for the ith sample times the width of the net – 0.085km | a\_i <- raw$effort |
| 11) Including vessel effects as overdispersion | There is only a single vessel for each survey; therefore, 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 |  |

Table 3. Model selection criteria for the different model combinations.

Figures

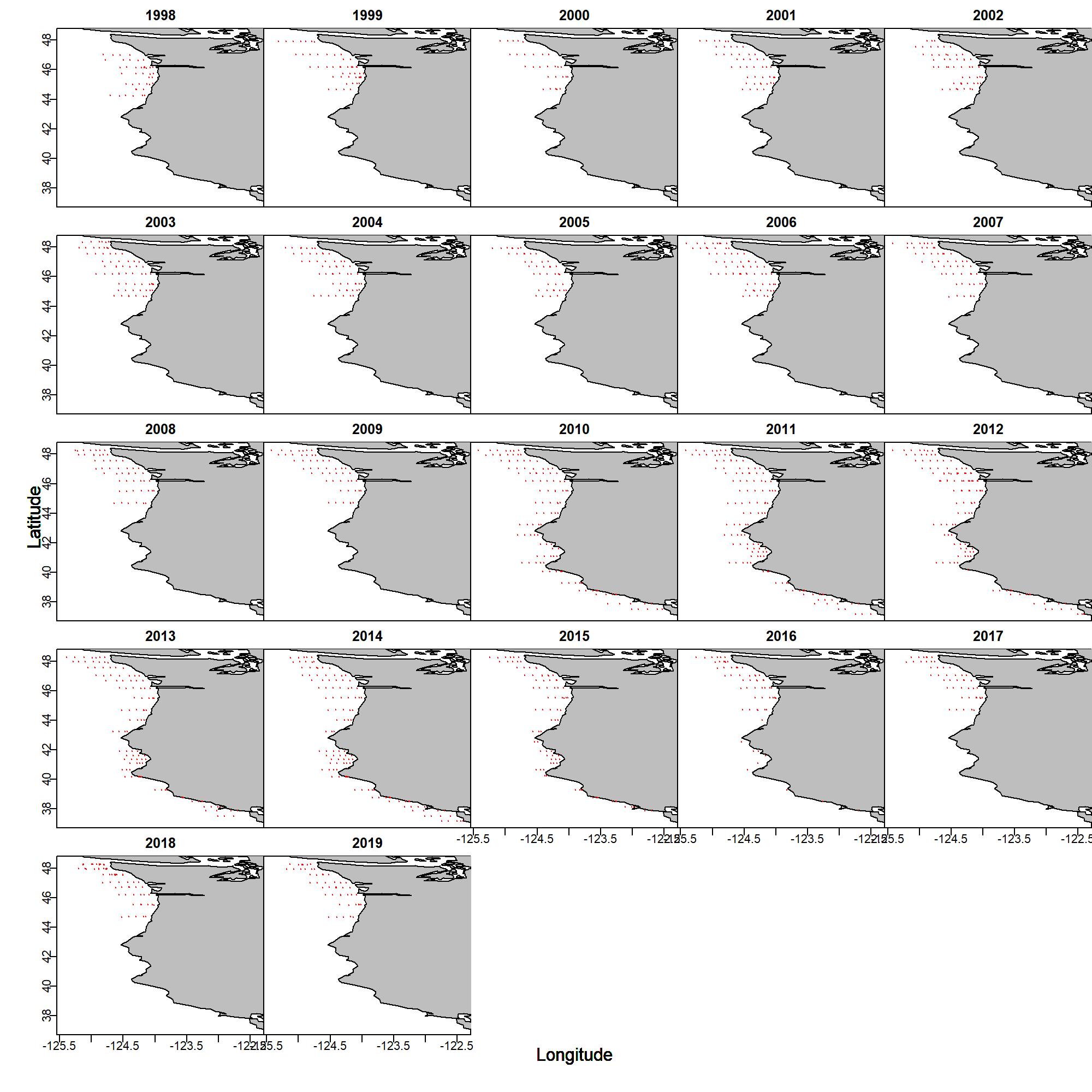


Figure 1. Location of samples for the 22 year data set. (We should probably color code by survey).

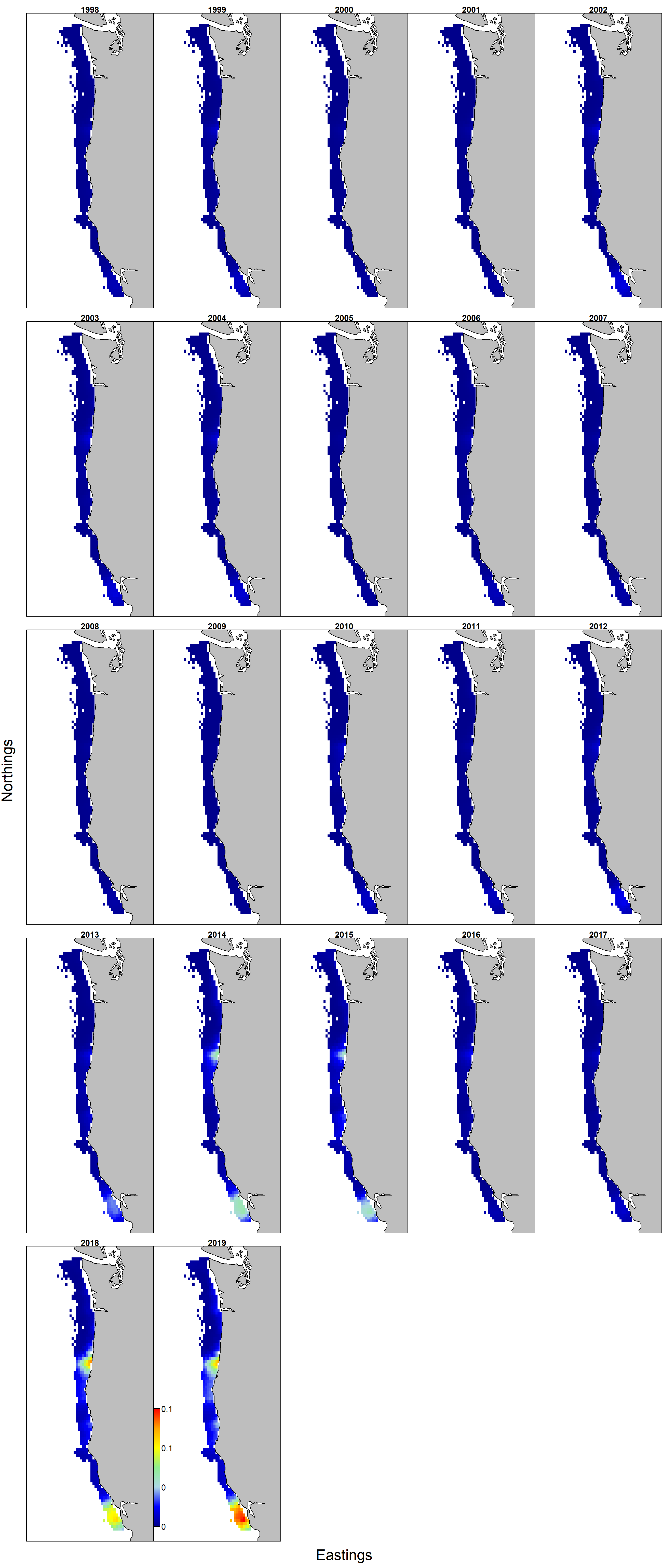


Figure 2. Estimated encounter probabilities for small () and () large squid collected during fisheries independent surveys by the NWFSC and SWFSC between 1998 and 2019

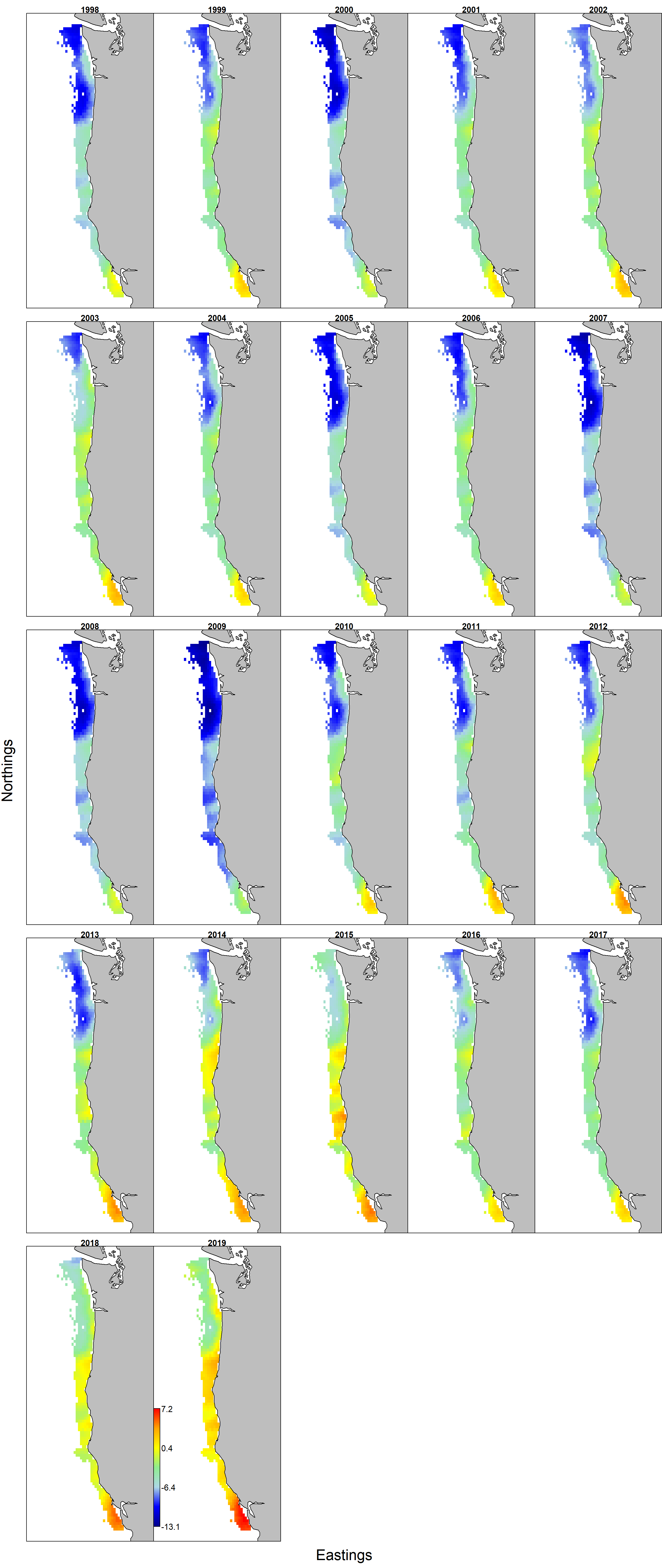


Figure 3. Estimated log-transformed densities for small () and () large squid collected during fisheries independent surveys by the NWFSC and SWFSC between 1998 and 2019.

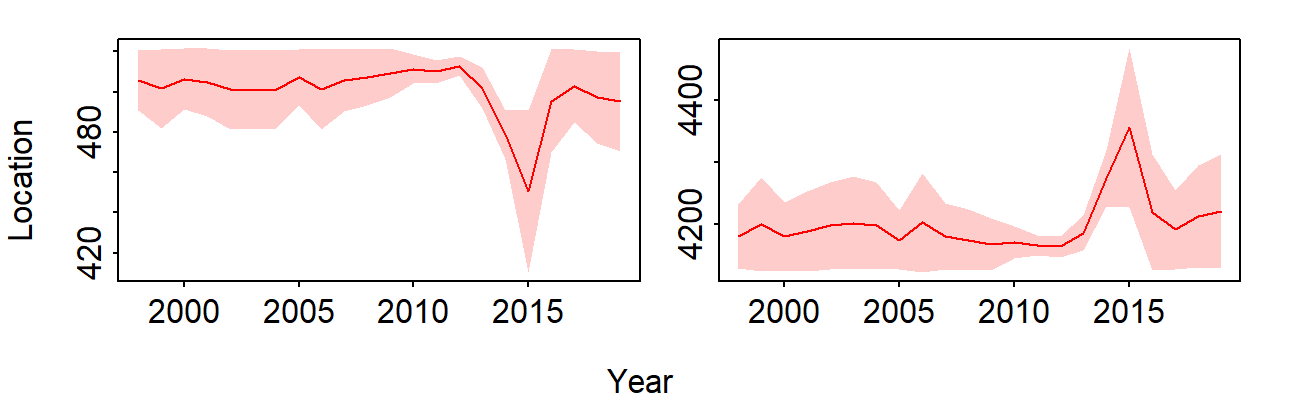


Figure 4. Estimated temporal easting (left panel, larger values more easterly) and northing (right panel, larger values more northerly) shifts in the center of gravity of the distribution for small and large squid collected during fisheries independent surveys conducted by the NWFSC and SWFSC between 1998 and 2019.

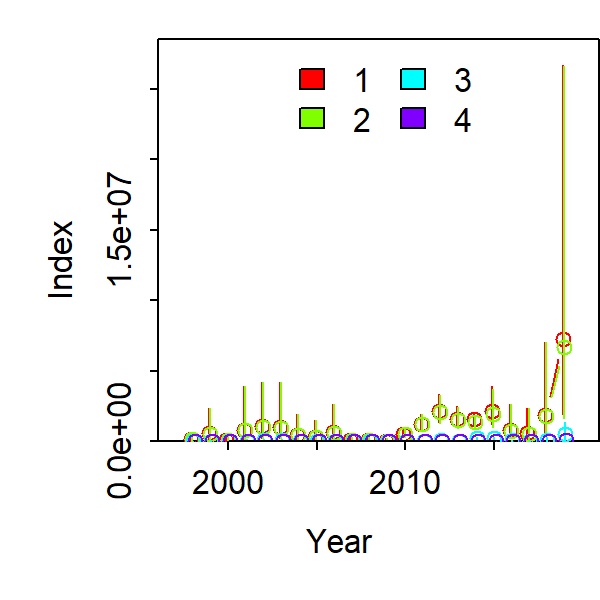


Figure 5. Estimated indices of abundance for the small and large squid collected during fisheries independent surveys conducted by the NWFSC and SWFSC between 1998 and 2019.

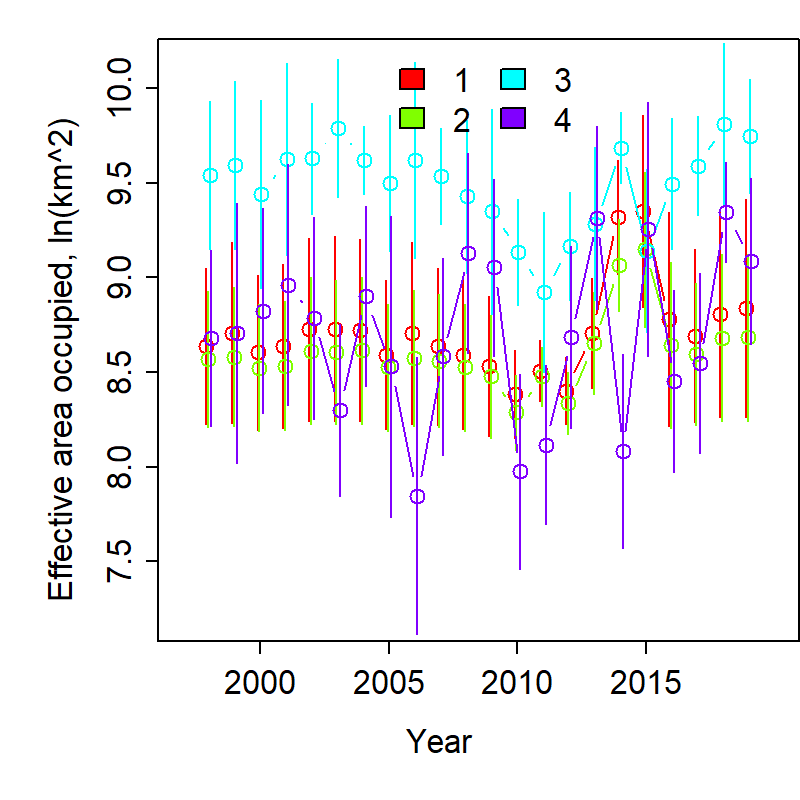


Figure 6. Range occupancy for small and large squid in the waters surveyed by the NWFSC and SWFSC from 1998 to 2019.