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 the Southwest Fisheries Science Center (SWFSC) from 2010 to 2016 (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 catches were less than a few hundred, otherwise the total catch of squid was extrapolated from sample weights or volumes. 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, the vessel, and the 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. We could compare the spatiotemporal synchrony for the NWFSC and SWFSC survey catches; however, with no spatial overlap, and only a small number of years of data 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. See Table 1 for a description of the subscripts, variables, and parameters of the model. The geostatistical model was delta-generalized linear model (delta-glmm) consisting of two parts (), the probability of encountering squid during a survey (i.e., encounter rate), and the probability of positive catches if squid were encountered

Equation .

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, 0.028 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 encounter rate and positive catches. For the positive catches observation model, we did not have true counts of the squid (i.e., the number of squid were often extrapolations based on weights or volumes) so we chose the gamma distribution (i.e., , where is equal to the coefficient of variation for the observations). For the encounter probability, we assumed a Bernoulli distribution ().

The delta-glmm model for the encounter probability (Equation 2) and positive catches (Equation 3) included a linear combination of estimators with i) intercepts, ii) covariates, and iii) spatial and spatiotemporal processes,

Equation .

Equation .

Based on an initial analysis of the size distribution data (Supplemental Figure S 1), we found little consistency in distinct length modes that would support a dynamic factor analysis of squid cohorts. Market squid grow quickly, spawn continuously, and only live one year (citation), making it unlikely that an annual survey would detect cohort differences. The vectors of intercept parameters, and , are the annual changes in the encounter rate and positive catches, respectively, for the ith observation, and they maybe estimated as either fixed effects or random effects based on auto-regressive process (AR1). For example, using the intercepts for the encounter rate, the AR1 process for intercepts is defined as , where is the correlation between time steps, and is the variance. The positive catch intercepts would described by a similar process but with different subscripts for the correlation and variance parameters.

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 rate 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 axes 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 (2019) fifteen step decision tree when implementing the spatiotemporal model in the VAST package when exploring various model structures. See Supplemental Table S 1 for a description of model inputs that pertain to the model with the best fit to the data. 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. We used AIC to compare models combinations of catchability coefficients, spatial and spatiotemporal processes, and temporal correlation for the intercepts parameters and spatiotemporal processes (Akaike 1974). Finally, for the top three model chosen based on AIC, we used a k-fold cross-validation with 10 folds to determine how well each model performed on unknown “out-of-bag” datasets. Specially, we compared the sum of the negative log-likelihood for the “out-of-bag” datasets across all three models, where the lower the total negative log-likelihood is, the better a model predicts the unobserved data.

**Results**

We were able to fit geostatistical models to fisheries independent squid catches from 1998 and 2019, and the model with the most parsimonious fit to the data included spatial and spatiotemporal variation for the encounter rate and density of squid, suggesting a large amount of geospatial variability among years for the two observation processes (Table 2). The amount of geospatial variability is evident in the ‘hotspots’ of encounter rates (Figure 2) and squid density (Figure 3) among years; however, both the encounter rates and densities show a trend toward higher encounter rates and catches in the southern regions of our study domain.

Results from the model suggest that the center of gravity (COG) for the survey catches has been relatively stable over the last twenty two years except for a period between 2014 and 2016, when the COG shifted up to 150 km north and 30 km west (Figure 4). Those shifts in the COG are commensurate with increased catches in northern California and the Oregon strata (check on this) and the manifestation of the warm blob that existed in the northeast Pacific from 2014 to 2017. Model results also suggest a range expansion since 2010 for the California, Oregon and Washington strata as demonstrated by an increase the effective area occupied (EAO, i.e., the index of abundance divided by the average density; Figure 5). Among the strata, the range expansion since 2010 has varied between 2,500 km2 for California, to ~7,500 km2 for Oregon and Washington.

After accounting for the geospatial differences in the distribution of the squid, the estimated index of abundance has increased across all strata since 1998 (Figure 6, panel A), with index catches of squid/km2 during the 2019 survey approaching 55,000, 6,800 and 1,300 for California, Oregon and Washington, respectively. While the survey catches from California dominated the coastwide aggregate catch (Figure 6, panel A), the relative increase in the indices has been similar across all strata, with a 59-fold, 116-fold, and 69-fold increase in index catches from 1998 to 2019 for California, Oregon and Washington, respectively (Figure 6; panel B). I would like

We used several diagnostic tools to assess the model fit the data. Visual inspection of the quantile plots for both the encounter rates and positive catches (Supplemental Figure S 2 and Supplemental Figure S 3, respectively) provide support that the distributions of the data and model predictions are similar. Second, plots of the Pearson residuals for the encounter rates and positive catches show no indications of correlation deviates in either space our time that would bias the parameter estimates or derived variables (Supplemental Figure S 4 and Supplemental Figure S 5, respectively). Finally, we compared the predictive ability of the top three models, using a 10-fold cross validation. Using the sum on the negative log-likelihood for the 10-folds, we found the model with the lowest AIC (m=1) also had the highest predictive for the “out-of-bag” samples , while

**Discussion**

**Tables**

Table . 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 . The shows the relative difference of the Akaike information criteria compared to the model with the lowest AIC. The column headings refer to whether or not the catchability covariates or a random process was included in the model. A one (1) implies the covariates or process were estimated and zero (0) implies it was not. For the spatial and spatiotemporal columns, the first number refers to the encounter rate model and the second number refers to the density model. For the temporal process the first values refers to the correlation for the intercepts in both the encounter rate and density models, and the second number refers to the correlation of the intercepts for the catchability coefficients for both the encounter rate and density models. Because all of the top models included spatial and spatiotemporal random effects, we compare the eight models with or without covariates and different combinations of temporal correlation for the random processes.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Temporal correlation | Spatial | Spatiotemporal | Parameters | AIC |
| 1 | 0 1 | 1 1 | 1 1 | 476 | 0 |
| 2 | 0 0 | 1 1 | 1 1 | 472 | 4.1 |
| 3 | 0 1 | 0 1 | 1 1 | 468 | 6.3 |
| 4 |  |  |  |  |  |
| 5 |  |  |  |  |  |
| 6 |  |  |  |  |  |
| 7 |  |  |  |  |  |
| 8 |  |  |  |  |  |

**Figures**

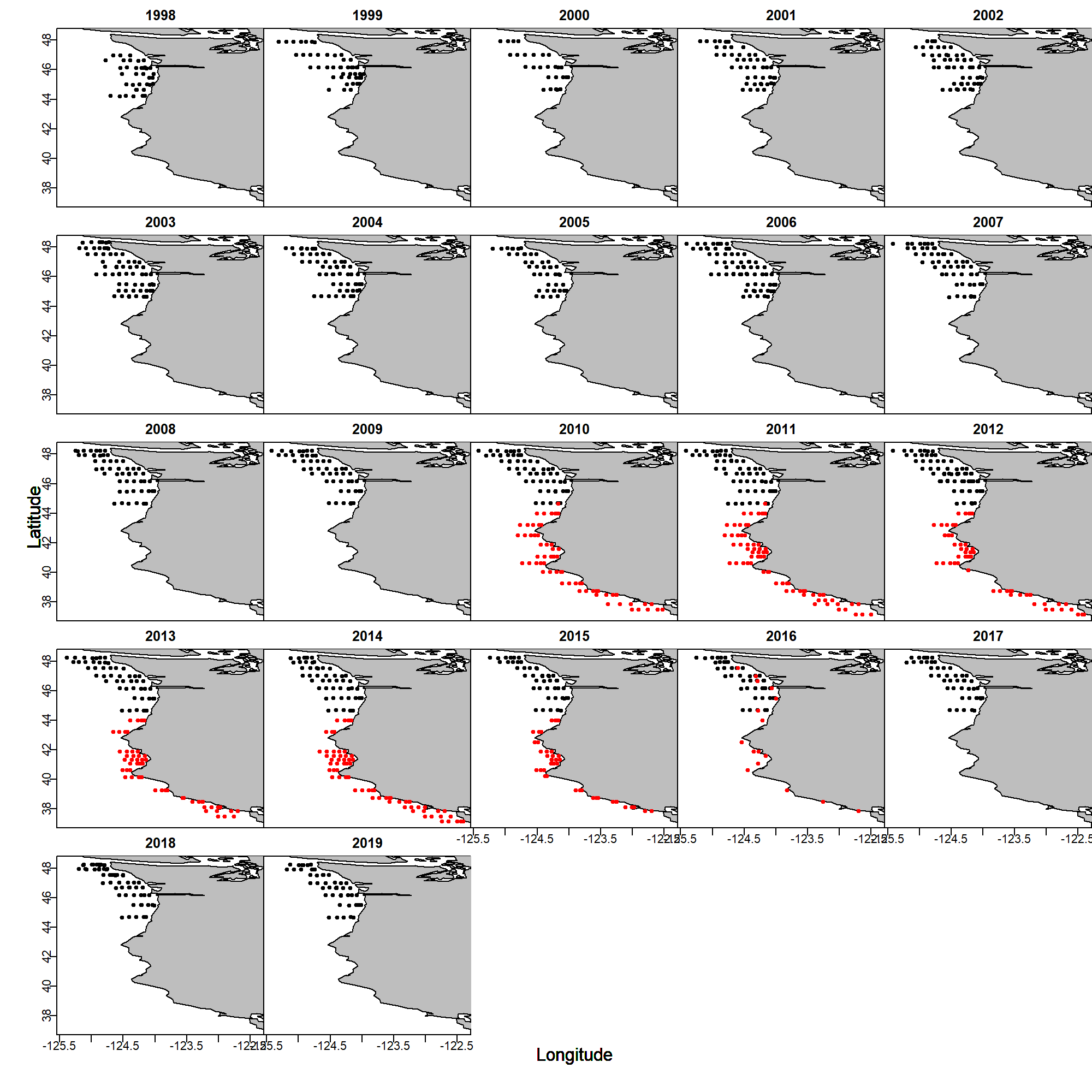


Figure . Location of samples from the NWFSC (black) and SWFSC (red) surveys from 1998 to 2019.

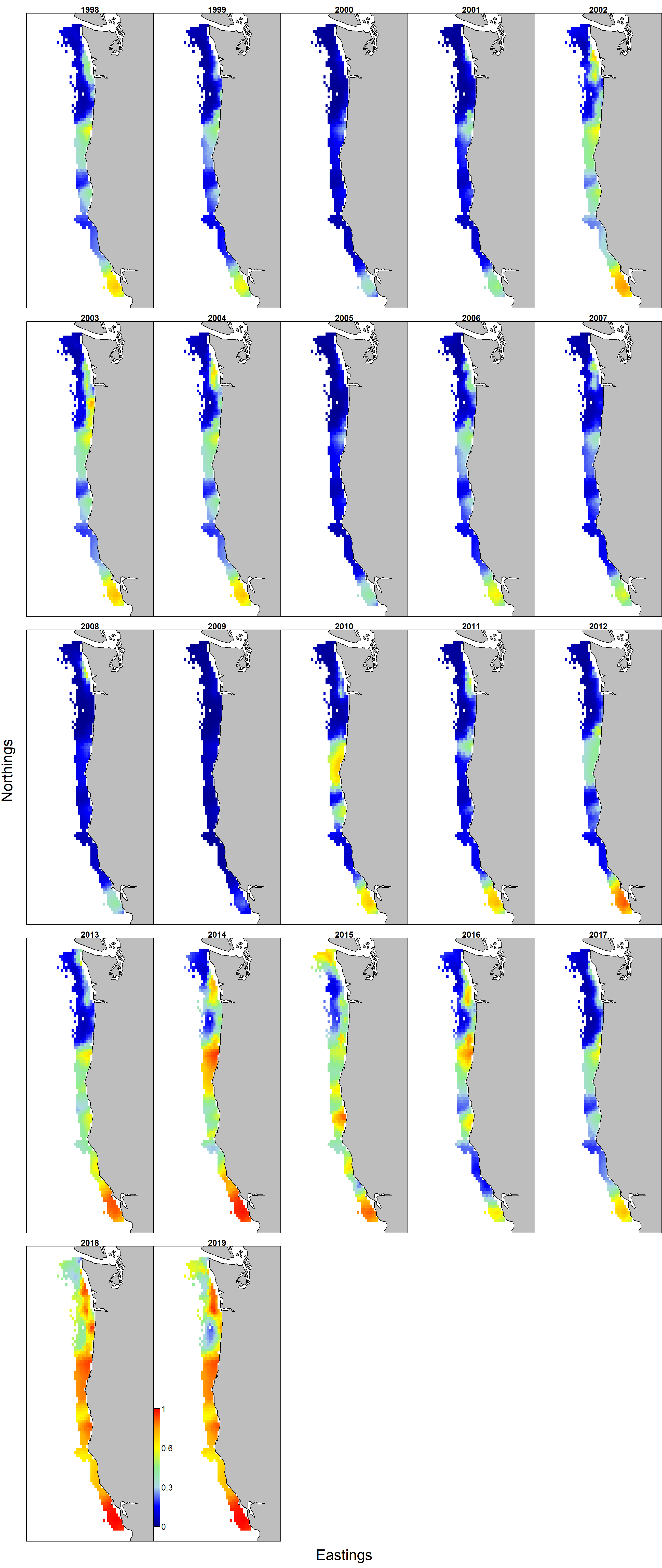


Figure . Estimated encounter rates for market squid collected during fisheries independent surveys by the NWFSC and SWFSC between 1998 and 2019

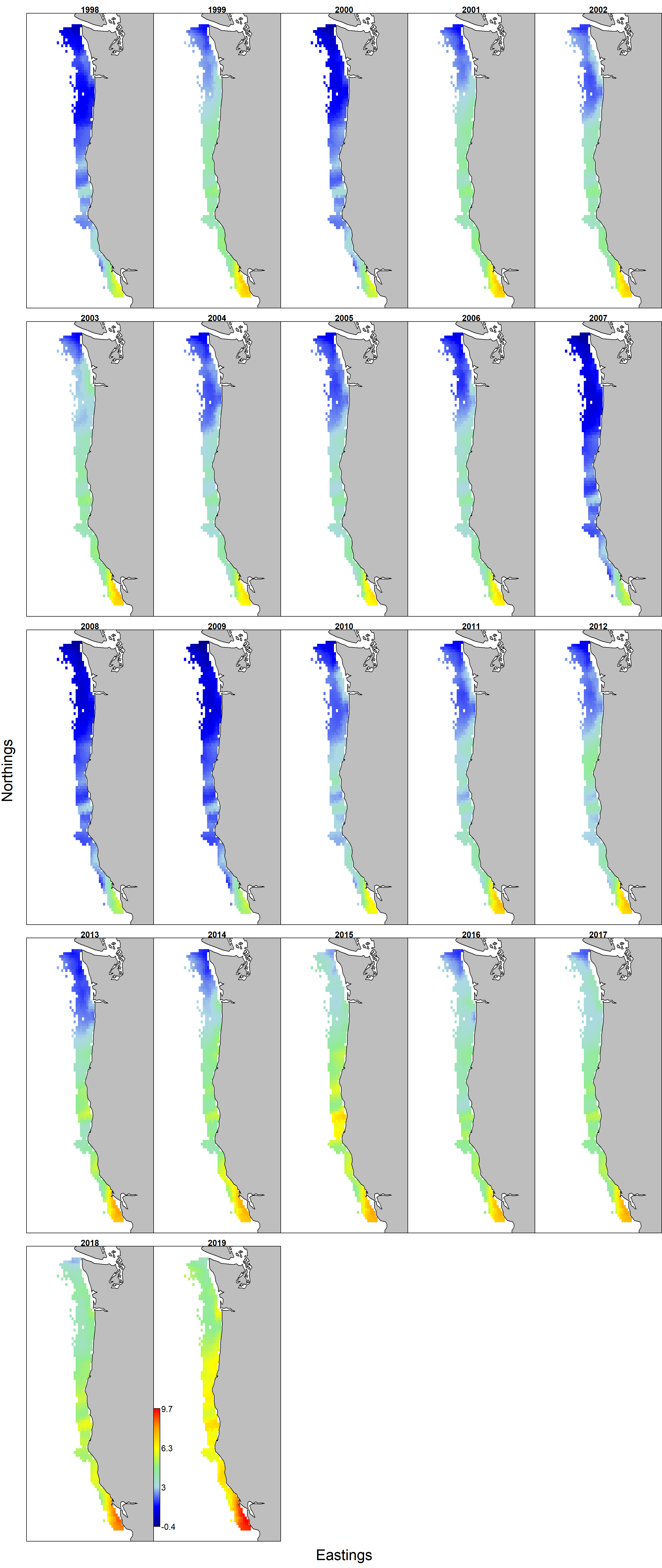


Figure . Estimated log-transformed densities for market squid collected during fisheries independent surveys by the NWFSC and SWFSC between 1998 and 2019.

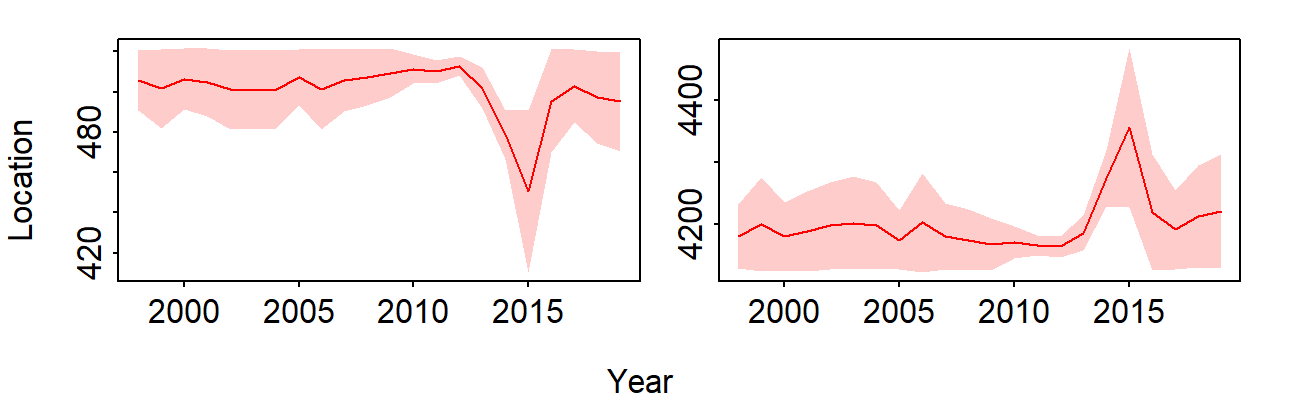


Figure . Estimated temporal easting (left panel, larger values more easterly) and northing (right panel, larger values more northerly) shifts in the center of gravity of the distribution for small and large squid collected during fisheries independent surveys conducted by the NWFSC and SWFSC between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

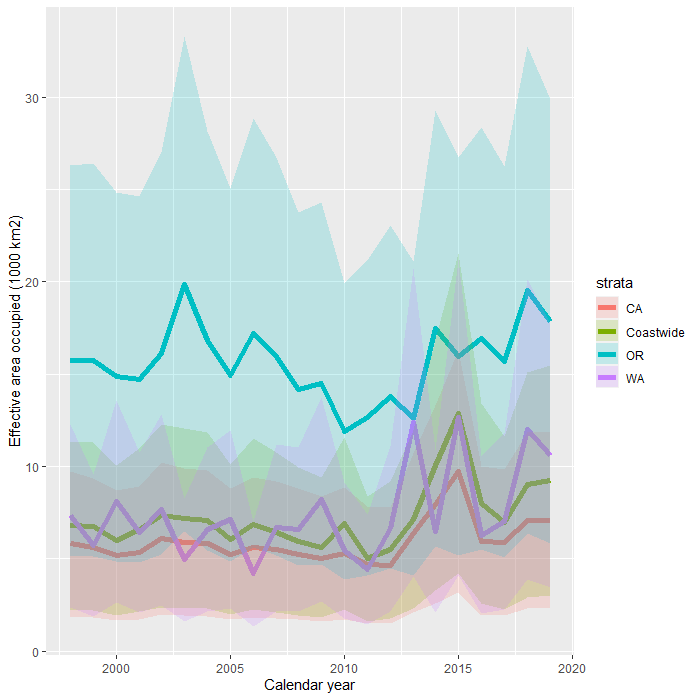


Figure . Effective area occupied and 50% credible intervals (1000 km2) for market squid in the waters surveyed by the NWFSC and SWFSC from 1998 to 2019. The effective area occupied measures the area needed to contain a population given its average population density (numbers km−2).

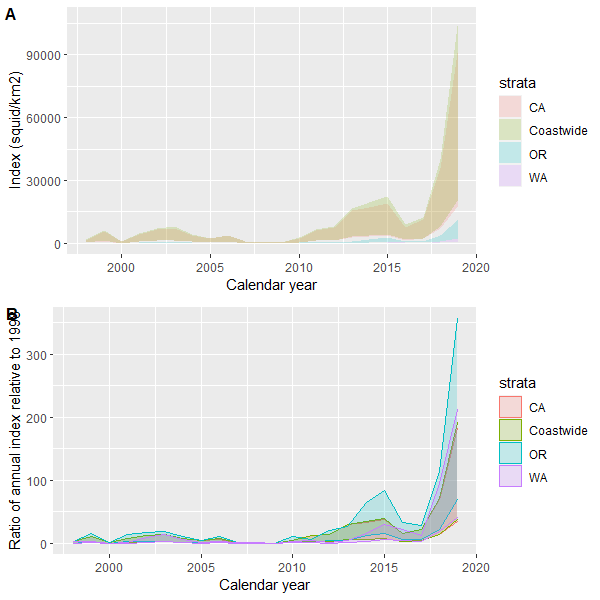
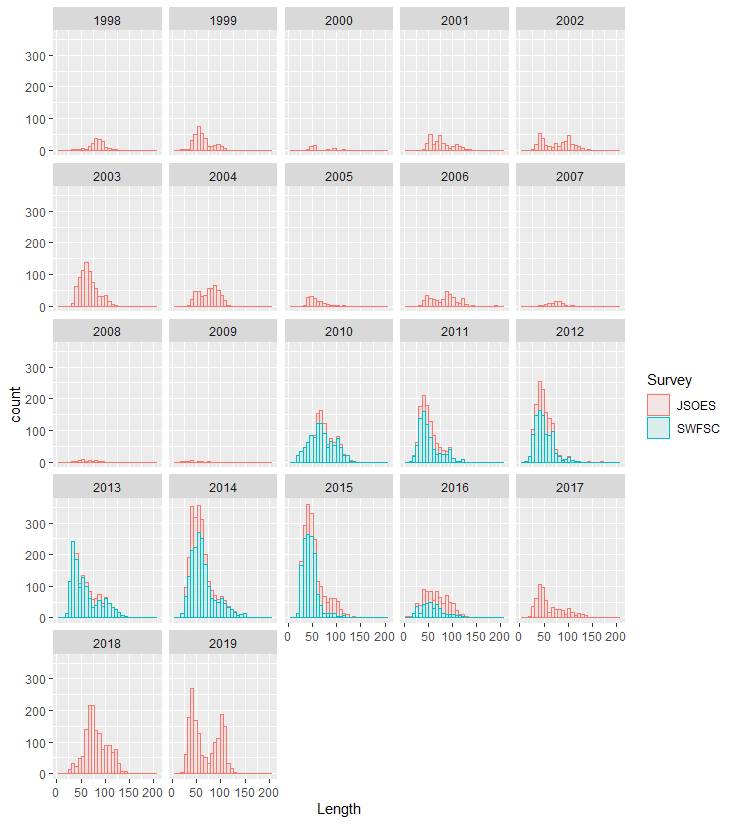


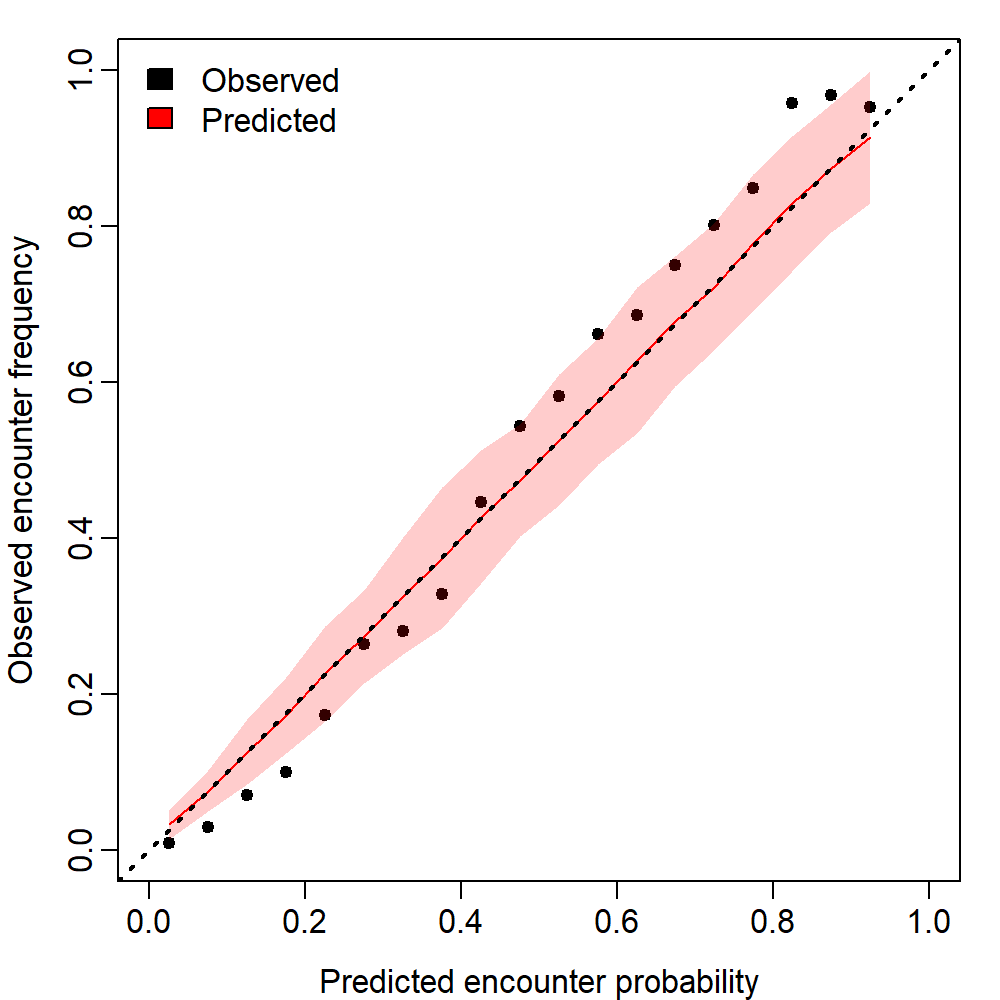
Figure . Estimated 50% credible intervals for the index (catch/km2; panel A) and percent change in the index (panel B) of market squid collected during fisheries independent surveys conducted by the NWFSC and SWFSC between 1998 and 2019 in the nearshore waters of California, Oregon and Washington.

Supplemental Table S . Decision tree for the VAST package based on Thorson 2019 and the model input for our model with the lowest AIC.

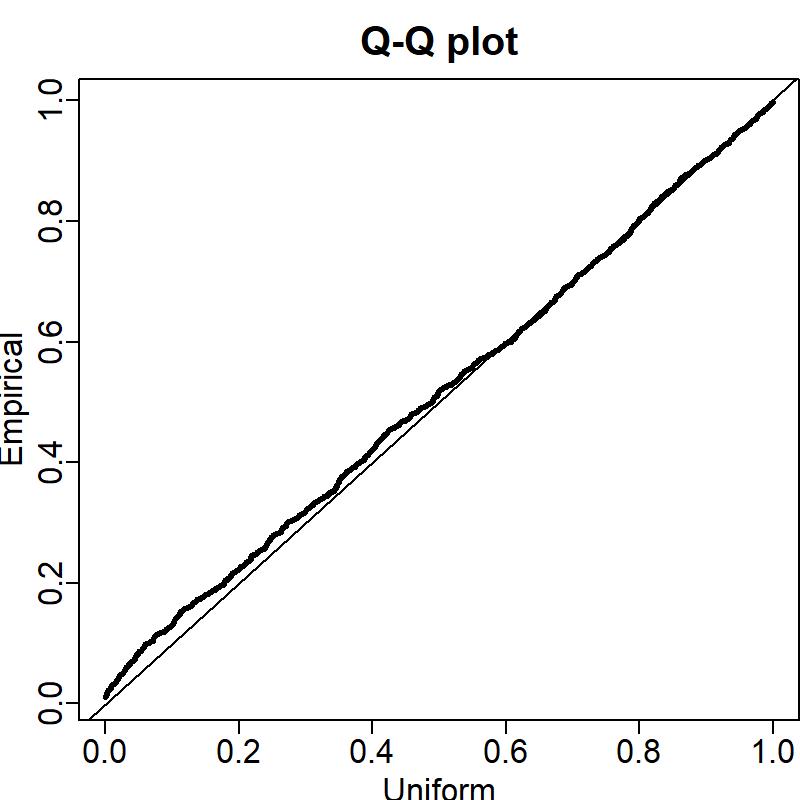
|  |  |  |
| --- | --- | --- |
| Decision | Description | VAST |
| 1) Spatial domain used when calculating derived quantities | The west coast of the US from San Francisco Bay to the northern tip of Washington State. We divide the coast into four strata, with ‘coastwide’ as an aggregate for the three 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 | 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" = 1, "Epsilon1" = 1, "Omega2" = 1, "Epsilon2" = 1) |
| 5) Choosing the spatial smoother and resolution | We used a “mesh” grid, with 175 nodes, and assume geometric anisotropy – east/west and north/south deviates are not symmetric. | Mesh.Method <- "Mesh"  n\_x <- 175  Aniso <- TRUE |
| 6) Choosing the number of spatial and spatio-temporal factors | Because our model does not include any size factors, the FieldConfig parameters remains the same. | FieldConfig <-c(Omega1 =1, Epsilon1 =1, Omega2 =1, Epsilon2 =1) |
| 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 |  |



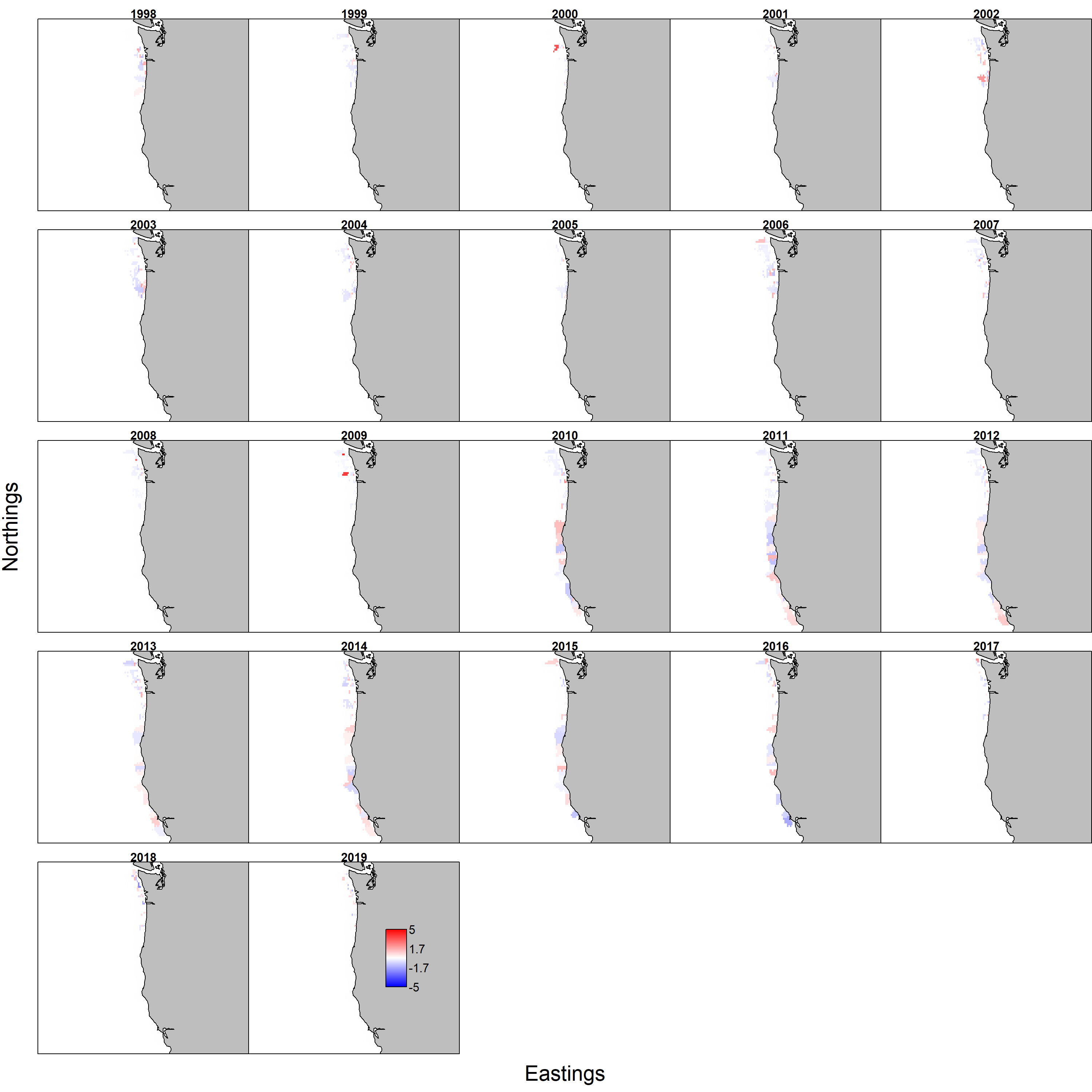
Supplemental Figure S . Supplemental Figure S1. Distribution of length classes of market squid captured between 1998 and 2019 for the NWFSC and SWFSC surveys.



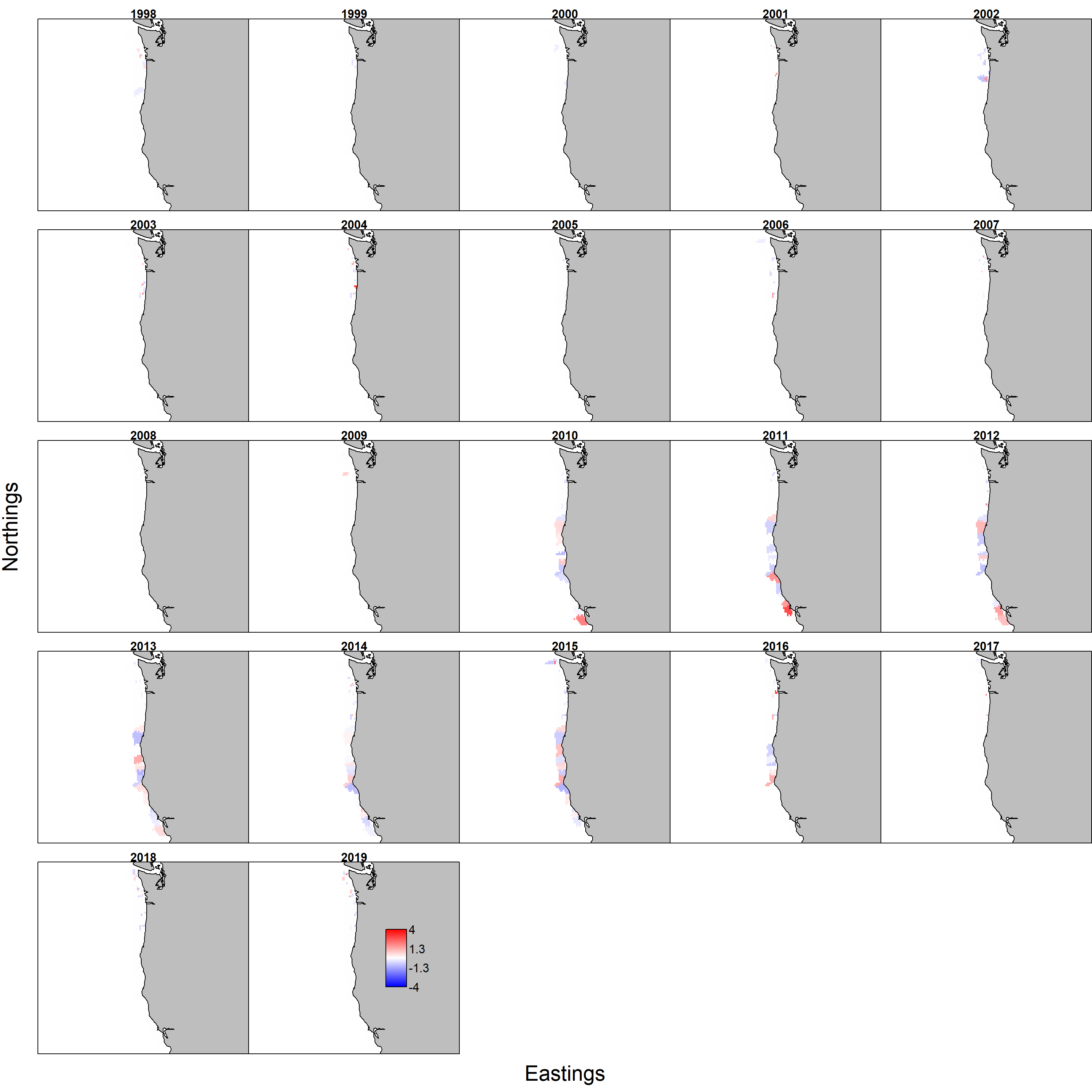
Supplemental Figure S . Quantile plot comparing the observed and predicted (with 95% predictive probability) distribution for the encounter probability.



Supplemental Figure S . Quantile plot comparing the cumulative distribution of empirical observation and the model predictions.



Supplemental Figure S . Pearson residuals for the encounter rates of market squid across the spatial domain from 1998 to 2019.



Supplemental Figure S . Pearson residuals for the positive catches of the market squid across the spatial domain from 1998 to 2019.