Appendix C: Supplmentary materials for: Canaries of the Arctic: the collapse of eastern Bering Sea snow crab

Cody Szuwalski (and others to come)

Supplementary materials

Methods overview

We used an integrated population model to estimate variation in mortality over time for snow crab in the eastern Bering Sea and generalized additive models (GAMs) to relate the estimated variation in mortality to potential stressors in the environment. The population dynamics model was fit to abundance and size composition data from the National Marine Fisheries Service (NMFS) summer bottom trawl survey in the eastern Bering Sea shelf to estimate total mortality by maturity state and year for male snow crab. We then developed indices for temperature occupied, disease prevalance, cannibalism, and density dependent effects from the NMFS survey to test as covariates in GAMs. Indices for fishery related effects were collated from fisheries statistics from the Alaska Department of Fish and Wildlife and also included in the GAMs. Below we describe each of these components, discuss the rationale behind our modeling decisions, and provide sensitivities and simulation tests of our models.

Population dynamics model

The population dynamics model presented here incorporates the best available information on relevant population processes to estimate total mortality for male snow crab on the eastern Bering Sea shelf and is similar in structure to the model used to assess eastern Bering Sea snow crab for management (Szuwalski, 2021). The model tracked numbers of male crab at size at maturity state over time with size bins ranging from 30-95 mm carapace width with 5 mm bin widths. Only male crab were modeled because male and female crab appear to have somewhat different dynamics and the male crab in the modeled size range are better selected by the survey gear (Szuwalski, 2021). Snow crab are sexually dimorphic, with male snow crab growing to nearly twice the size of females, which accounts for the better selection in the survey. Only crab smaller than 95 mm were modeled for two reasons: 1) to attempt to isolate the effect of the directed fishery (crabs of >101 mm carapace width are targeted in the fishery; discussed further below) and 2) almost all of the crab that disappeared since 2018 are in this size range. The population dynamics model operates on a half year time step, starting in July at the time of the NMFS survey. The fishery is assumed to occur in February. Total mortality (Z) is estimated by year (y) and maturity state (m). Other estimated parameters include the initial numbers at size by maturity state, yearly log recruitments, a vector of scalars that determine the proportions of estimated recruitment split into the first two size bins, and a variance component for the penalty on total mortality. Parameters determining growth, maturity, and survey selectivity were estimated outside of the model and specified when estimating mortality and catchability. Mortality is the only population process that occurs in the first half of a given year:

$$N_{t=y+0.5,s,m} = N_{t=y,s,m} e^{-Z_{t,s,m}/2} \tag{1}$$

Growth occurs at the beginning of the second half of the year for immature crab and is represented in the model by multiplying the vector of immature crab at size by a size-transition matrix $X_{s,s'}$ that defines the

size to which crab grow given an initial size. Snow crab are observed to undergo a 'terminal molt' to maturity after which growth ceases (Dawe et al., 1991). Accordingly, all immature crab are assumed to molt and no mature crab molt in our model. The newly molted crab are assigned to a maturity state based on observed ogives of the proportion of mature new shell males by size calculated from chelae height measured in the NMFS survey data (Otto, 1998), which varies over time ($\rho_{y,s}$; Figure 4). The average probability of having undergone terminal molt is used in years during which data were not collected. This process results in two temporary vectors of numbers at size:

$$n_{t=y+0.5,s,m=1} = \rho_{y,s} X_{s,s'} N_{t=y+0.5,s,m=1} \tag{2}$$

$$n_{t=y+0.5,s,m=2} = (1-\rho_{y,s}) X_{s,s'} N_{t=y+0.5,s,m=2} \tag{3}$$

The size transition matrix $X_{s,s'}$ was constructed using growth increment data collected over several years (see Szuwalski, 2021 for a summary) to estimate a linear relationship pre- and post-molt carapace width (Figure 5), ($\hat{W}^{pre}_{s,w}$ and $\hat{W}^{post}_{s,w}$, respectively) and the variability around that relationship was characterized by a discretized and renormalized normal distribution with a size-varying standard deviation, $\mathrm{Y}_{s,w,w'}$ (Figure 5).

$$X_{s,w,w'} = \frac{Y_{s,w,w'}}{\sum_{w'} Y_{s,w,w'}} \tag{4}$$

$$Y_{s,w,w'} = (\Delta_{w,w'})^{\frac{\hat{L_{s,w}} - (\bar{W}_w - 2.5)}{\beta_s}} \tag{5}$$

$$\hat{L}^{post}_{s,w} = \alpha_s + \beta_{s,1} hat W^{pre}_{s,w} \tag{6}$$

$$\Delta_{w,w'} = \bar{L}_{w'} + 2.5 - W_w \tag{7}$$

It is important to note that crab can 'outgrow' this model, which is represented by the pre-molt-carapace widths (e.g. 87.5 and 92.5 mm carapace width in Figure 5) that have low probability of molting to any of the sizes that are included in the population dynamics model.

Recruitment by year, τ_y, was estimated as a vector in log space and added to the first two size of classes of immature crab based on another estimated vector δ_y that determines the proportion allocated to each size bin.

$$n_{t=y+0.5,s=1,m=1} = n_{t=y+0.5,s,m=1} + \delta_y e^{\tau}_y \tag{8}$$

$$n_{t=y+0.5,s=2,m=1} = n_{t=y+0.5,s,m=1} + (1-\delta_y) e^{\tau}_y \tag{9}$$

Finally, the last half of the year of mortality is applied to the population after growth, maturity, and recruitment occurs. Note that this allows a crab to experience two different mortalities within a given year as it undergoes terminal molt.

$$N_{t=y+1,s,m=1} = n_{t=y+0.5,s,m=1} e^{-Z_{t,s,m}/2} \tag{10}$$

$$N_{t=y+1,s,m=2} = (N_{t=y+0.5,s,m=2} + n_{t=y+0.5,s,m=2}) e^{-Z_{t,s,m}/2} \tag{11}$$

Survey selectivity

The observed numbers of crab at size by year in the NMFS survey reflect the ability of the trawl gear to capture the crab, also known as 'selectivity'. The selectivity of trawl gear can change according to size, and consequently needs to be accounted for in the population dynamics model when fitting to the survey data. Values for survey selectivity at size were specified using data from experimental trawls conducted by the Bering Sea Fisheries Research Foundation in collaboration with the NMFS summer survey. The experimental trawls were performed at the same time and location as the NMFS summer survey tows to evaluate the efficiency of the NMFS survey trawl gear at capturing snow crab (Somerton et al., 2013). The nephrops gear used by the BSFRF was assumed to capture all crab in its path given strong bottom contact. The resulting area-swept estimates of numbers of crab at size from the BSFRF and NMFS surveys ($\hat{N}_{y,s,NMFS}$ and $\hat{N}_{y,s,NMFS}$, respectively) can be used to infer the selectivity of the NMFS gear in year y as:

$$S_{y,NMFS} = \frac{\hat{N_{y,s,NMFS}}}{\hat{N}_{y,s,BSFRF}} \tag{12}$$

The experimental trawls captured snow crab in the years 2010, 2011, 2016, 2017, and 2018, but the spatial foot print and sample sizes varied by year (Figure 6). The calculated selectivities by size and by year were fairly consistent for snow crab of carapace widths 40 - 95 mm, but the signal was less consistent for crab larger than ~100 mm carapace width (Figure 7). The selectivity of large crab determines the estimated scale of the population in a population dynamics models, but the information we have on selectivity of is poor and different assumptions about selectivity lead to very different inference about the stock (Szuwalski, 2021b). The lack of clear information on the scale of the population exploited by the fishery is one of the key reasons we used the range of sizes included in this model and excluded the directed fishery data from the analysis. A GAM was fit through the estimates of selectivity and the resulting estimates by size were directly specified in the population dynamics model.

'Catchability' represents the fraction of the population available to the survey gear (either as a result of spatial mis-match or the inability of the gear to come in contact with the animals as a result of burrowing or hiding in untrawlable habitat). The capability for modeling time-varying catchability was built into the model in the form of a vector of parameters equal to the length of the time series of data. When time-varying catchability was estimated, the yearly catchability parameters were used to scale the selectivity curve described above up or down.

Objective function

The objective function for the population dynamics model consists of likelihood components and penalty components that are summed and minimized in log space to estimate parameters within the model. Several data sources were fit to using the following likelihoods. Observed size composition data for immature and mature males were fit using multinomial likelihoods and were implemented in the form:

$$L_x = \lambda_x \sum_y N_{x,y} \sum_l p^{obs}_{x,y,l} ln(\hat{p}_{x,y,l}/p^{obs}_{x,y,l}) \tag{13}$$

L_x was the likelihood associated with data component x, where λ_x represented an optional additional weighting factor for the likelihood, $N_{x,y}$ was the sample sizes for the likelihood, $p^{obs}_{x,y,l}$ was the observed proportion in size bin l during year y for data component x, and $\hat{p}_{x,y,l}$ was the predicted proportion in size bin l during year y for data component x. Sample sizes were input as 100, which is the value currently used in the stock assessment (Szuwalski, 2021).

Observed indices of abundance for immature and mature males were fit with log normal likelihoods implemented in the form:

$$L_x = \lambda_x \sum_y \frac{(ln(\hat{I}_{x,y}) - ln(I_{x,y}))^2}{2(ln(CV_{x,y}^2 + 1))} \tag{14}$$

L_x was the contribution to the objective function of data component x, λ_x was any additional weighting applied to the component, $\hat{I}_{x,y}$ was the predicted value of quantity I from data component x during year y, $\mathrm{I}_{x,y}$ was the observed value of quantity I from data component x during year y and $\mathrm{CV}_{x,y}$ was the coefficient of variation for data component x during year y.

Penalties and priors

Smoothing penalties were placed on estimated vectors of deviations for immature and mature natural mortality and immature and mature catchability. A prior value of 0.27 is used for the average natural mortality based on assumed maximum age of 20 and Hamel's (2015) empirical analysis of life history correlates with natural mortality. The priors used for catchability were derived from the selectivity experiments described above. Penalties were implemented using normal likelihoods on the second differences of the vector. A separate normal prior was placed on the estimated mean value of immature and mature mortality in the form:

$$L_x = \lambda_x \sum_y \frac{((\hat{I}_{x,y}) - (I_{x,y}))^2}{CV_{x,y}^2} \tag{15}$$

L_x was the contribution to the objective function of data component x, λ_x was any additional weighting applied to the component, $\hat{I}_{x,y}$ was the predicted value of quantity I from data component x during year y, $\mathrm{I}_{x,y}$ was the observed value of quantity I from data component x during year y and $\mathrm{CV}_{x,y}$ was the coefficient of variation for data component x during year y.

Population dynamics model sensitivities

Modeling decisions are necessarily made in the process of writing population dynamics models and it is possible for these decisions to influence the outcome of an analysis. Within the context of our model, these decisions include what processes to allow to vary over time, the weights assigned to different data sources and portions of the objective function, which parameters to place priors or penalties on, and what those priors or penalties should be. We ran several sensitivity analyses to understand the implications of these modeling decisions on the outcome of our analysis.

Does allowing mortality or catchability to vary over time improve model fits?

Catchability and mortality are somewhat confounded within population dynamics models (Thompson, 1994). Fewer crab observed in a given year can be attributed to either crab dying or by crab moving out of the surveyed area either by walking out of the boundaries or burying themselves into the substrate. At the same time, it is also clear that catchability and mortality likely vary over time in reality in spite of the fact that they are often assumed to be time-invariant in population dynamics models (Johnson et al., 2014). Somerton et al. (2013) showed that catchability varied somewhat by substrate and depth for snow crab in the EBS. The spatial distribution of snow crab varies over time and substrate and depth vary over space, so it follows that catchability should also vary over time.

We started exploring the impacts of including time-variation in mortality and catchability on model output by fitting a model with no time-variation in mortality or catchability. Then we compared the output of this model to models that allow time-variation in mortality, catchability, and both processes simultaneously (Figure 8 & Figure 9). The model with no time-variation in mortality or catchability was able to capture the

general trend in immature and mature survey abundance solely through estimating variability in recruitment. Allowing time-variation in catchability improved the fits to immature survey abundances more than time-varying mortality, but time-variation in either process improved fits in a similar manner for mature survey abundances. Mature size composition data were fit similarly for all models, but immature size composition data were better fit by the models that allowed time-varying catchability (Figure 8). Part of the reason this difference in fits to immature size composition data occurs is the variability in the first several size bins resulting from the poor selectivity of the survey for small animals. Sometimes the peaks seen in larger size classes are reflected in the preceding years' data for the smallest size classes, sometimes those peaks are not reflected (compare Figure 10 to Figure 11). As a consequence, positive residuals occur in the smallest size classes when a pseudocohort is consistently seen in large size classes, but not observed in the smallest size bins (e.g. 1991 vs. 1992; 1997 vs 1998).

The model without time-variation in mortality or catchability explained 67% of the deviation in the abundance indices, time-varying mortality explained 77%, time-varying catchability explained 94%, and both processes varying explained 99% of the historical deviance. Model selection based on information criteria (e.g. AIC; Akaike, 1974) are often used to identify a model within a suite of models that most parsimoniously fits the data. Adding time-variation in natural mortality or catchability alone improved model fits parsimoniously (AIC of 3434.15 for base model vs. 1593.836 and 1321.486 for time-varying mortality and catchability, respectively). However, adding time-variation in both processes resulted in a higher AIC (1449.275) than implementing time-variation in catchability, owing to the large number of parameters estimated. While catchability and mortality are somewhat confounded, catchability is also confounded with other sorts of error (e.g. observation) and allowing a relatively unconstrained estimation of catchability over time resulted in over-fitting the data, the consequences of which will be seen in simulations below. Even with this paring of potential models, there are several assumptions that could influence the output of our models. The following sensitivities are aimed at exploring the impacts of those assumptions on model output.

How well can the model estimate mortality and selectivity with simulated data?

One of the most essential exercises to perform with a population dynamics model before using its output is to perform a 'self-test' in which data are simulated from the population dynamics model with appropriate error and then fit to with the model. The goal of this test is to determine whether or not a model can return the parameter values underlying the simulated data with the available quantity and quality of data. For our analysis, the ability of the model to estimate mortality and catchability are of particular interest because they are candidates for use as input into GAMs to attempt to link the estimates to environmental stressors. Recruitment is also of interest because of its confounding with the other processes.

Log-normal error was added to the true underlying abundance from the simulation model with three different coefficients of variation: 0.01, 0.10, and 0.30. Simulated data sets were generated 100 times under each observation error scenario and the population dynamics models were fit to them. Two population dynamics models were fit: one in which time-varying natural mortality was estimated and one in which time-varying natural mortality and time-varying catchability were estimated. Estimates of mortality were closer to the true underlying values than estimates of catchability (compare Figure 12 to Figure 13). Mature mortality was better estimated than immature mortality regardless of data quality or model configuration. The correlation between estimated and simulated mortality was 0.65 and 0.96 for immature and mature mortality for the 0.01 observation error scenarios, respectively. The ability of the models to estimate mortality became more similar as data quality decreased. Overall, the model was best able to estimate mature mortality and this is likely a consequence of its separation from estimated recruitment in time. In general, estimates of catchability for both maturity states were unreliable.

As a result of these simulation analyses, two modeling decisions arose. First, we used estimated variation in mortality from models that only estimate time-variation in mortality because the estimates of mortality from models that estimated time-variation in both mortality and catchabilty were less reliable. This precludes attempts to identify relationships between estimated catchability and environmental variables. Second, the inability of the model to capture the scale of the population (Figure 14) underscores the need to relate mortality to the environmental covariates outside of the model, rather than attempting to build them into

the model (similar to Dorn and Barnes, 2022). The covariates described below are indices of a particular environmental stressor, not absolute quantities that could provide scale to the model.

How do the assumptions about weighting and priors influence the estimated quantities?

Some aspects of the model that may influence the outcome of the fitting are specified by the user with no clear 'correct' value. These include the weights assigned to the size composition data, some priors placed on population processes, and the weights assigned to the smoothness penalties. We performed sensitivity analyses for these parameters to check how different specifications changed the fits to the data and the estimates of mortality and catchability. We input a range of values for the size composition weights (25, 50, 100), the prior on the mean natural mortality in log space (-1.6, -1.2, -0.8), the input standard deviation for the penalties on natural mortality (0.01, 0.1, 0.2) and the smoothness penalty on the estimated time series' of mortalities and catchabilities (0.001, 0.1, 0.5, 0.1).

Differences among sensitivity scenarios resulted in very small changes in the fits to the data (Figure 15), but larger changes in estimated mortalities and catchabilities (Figure 16). The smoothness penalty placed on mortality over time appeared to be the largest driver of changes in estimates of M and q, so we looked at a wider range of smoothness penalties (i.e. 0.001, 0.1, 0.25, 0.5, 1, 5, 10, 1000). Trajectories of mortalities were roughly preserved across this range. The prior on mean natural mortality predictably scaled the estimated time series up or down. The best available information suggests natural mortality should be approximately 0.27 given an assumed (but based on a range of studies; see Szuwalski, 2021 for a summary) maximum age of 20 years for wild snow crab. Based on these analyses, we elected to use small smoothing penalties because there is no evidence to suggest that mortality should be particularly smooth from year to year. These analyses also underscore the fact that the scale of the population is difficult to estimate with the data available and the need to relate mortality to the environmental covariates outside of the population dynamics model. This likely comes from the fact that recruitment and immature mortality are confounded (i.e. fewer immature crab in a given year can be because of increased immature mortality or because of lower recruitment) and the lack of data (like removals) given in an absolute (rather than relative) metric.

Covariate construction

A wide range of factors could potentially influence mortality of snow crab on the eastern Bering Sea shelf, including temperature, predation, disease, cannibalism, and fisheries effects. The NMFS summer trawl survey provides a rich spatio-temporal data set to develop time series of temperature occupied, predation, disease, and cannibalism. The fisheries-dependent observer data provide spatio-temporal information on bycatch.

Currently, estimating spatially-explicit, time-varying mortality is not computationally feasible, nor are data on movement available to inform such a model. Consequently, our analysis aggregates the spatial data for snow crab into time-series. The end goal is to use these time-series in predictive models to identify relationships between estimated mortality and stressors, so attention has to be paid to creating appropriate comparisons. For example, a predation index needs to consider not only the total consumption of crab by cod, but also the total number of crab in the ocean of the size that can be consumed by cod to be comparable to changes in mortality rates (discussed more below).

Another important point for consideration in covariate construction is the estimation of mortality by maturity state. Snow crab in the EBS undergo an ontogenetic migration in which juvenile crab settle on the northeast portion of the shelf after their pelagic phase, then migrate southwest into deeper and (usually) warmer waters (Ernst et al., 2005; Parada et al., 2010). This means that the conditions and stressors felt by immature crab can be different than the stressors felt by mature crab. To address this issue, the spatial data sets for temperature, disease, and cannibalism were split based on the size above which half of the population was mature. The size at which more than half of the population is mature changes by year, depending on recruitment dynamics and other demographic processes (Figure 17). After the survey data were split at the 50% at maturity size, time series of maturity-specific environmental stressors (Figure 18) were created as described below.

Temperature

Temperature is one of the key physical variables that structures the benthic ecosystem of the EBS (Mueter and Litzow, 2088). The cold pool, a mass of water <2 degrees Celsius, acts as a barrier to species interaction based on temperature preferences of different species. Snow crab are a stenothermic species, preferring cold water and juvenile snow crab in particular are rarely found outside of the cold pool (Dionne, 2003). The cold pool is directly related to the winter ice extent in the Bering Sea and has varied dramatically over time as the ecosystem moves between cool and warm stanzas (e.g. 2006-2010 vs. 2014-2019; Figure 1b of the main text and Figure 19). As the cold pool changes from year to year, so does the spatial distribution of snow crab (Figure 20). The ontogenetic migration of snow crab results in crab of different sizes and maturity states experiencing different temperatures in a given year (Figure 21). The 'temperature occupied' for different sizes of crab by year $T_{s,y}$ was calculated here as an average of the observed bottom temperatures at the stations at which crab of a given size were captured t_i, weighted by the area-swept density of crab at a given station d_i:

$$T_{s,y} = \frac{\sum_i d_i t_i}{\sum_i d_i} \tag{16}$$

The resulting time series of temperatures occupied by size were then split by maturity state by identifying a cutoff beyond which half of the population was mature and aggregating the temperatures above and below the cutoff to represent immature and mature temperature occupied (Figure 22).

Predation

Pacific cod (*Gadus macrocephalus*) are the largest predator of snow crab based on stomach content data collected in the NMFS summer bottom trawl survey (Long and Livingston, 1998). Immature crab under the size of 55 mm carapace width are the primary sizes consumed by cod in the Bering Sea (Burgos et al., 2010). Changes in the cold pool have altered the interaction between snow crab and Pacific cod over time. Decreases in the size of the cold pool coincide with more northerly positions of the centroids of abundance of cod (e.g. 2003 and 2018-2019; Figure 23 & Figure 24). This increased interaction coincided with increased numbers of crab consumed by cod in the last several years (Figure 25). However, this period of time also coincided with the appearance of the largest pseudo-cohort of snow crab ever seen in the Bering Sea. Given the generalist nature of Pacific cod, one would expect to see an increase in the amount of crab consumed by cod during this period of time even if there weren't differences in the interactions between the species as a result of changes in the cold pool. Further, a large fraction of the missing crab from the recent collapse were not of the sizes typically eaten by cod (Figure 26). To evaluate the possibility cod consumption has influenced the mortality of snow crab over time, the relative impact of consumption with respect to the population size must be considered. Consequently, predation indices were calculated for mature and immature animals by year $P_{m,y}$ by calculating the ratio of the extrapolated biomass of crab consumed by cod to the biomass of the estimated numbers of crab by maturity state, $N_{y,m,s} * w_s$:

$$P_{m,y} = \frac{cod_{y,m}}{\sum_s N_{y,m,s} * w_s} \tag{17}$$

The exact amount of crab eaten cannot be calculated from the diet data because the diet data are a snapshot of the consumption at one point during the year and consumption would be expected to change with spatial overlap and temperature-driven changes in metabolism occurring throughout the year. Consequently, removals due to predation cannot be directly incorporated into the model as fishery removals might be. However, the predation index developed here represents the best available information on the relative impact of cod predation on snow crab mortality for use in correlative models like the GAMs below.

Disease

Bitter crab syndrome is a fatal disease in snow crab caused by a parasitic dinoflagellate. The presence of disease is recorded in the NMFS summer trawl survey data for the subset of crab that are individually measured based on a visual inspection. Portions of the shells of diseased crab present as a milky white, which is different from their usual more translucent state. The spatial distribution of bitter crab disease is predominantly on the northeastern shelf where smaller immature animals are found (Figure 27). For this analysis, disease prevalence was calculated simply as the number of infected individuals identified in the survey divided by the total number of individuals caught in the survey (Figure 18).

Cannibalism

Cannibalism has been proposed as a potential driver of the dynamics of snow crab in eastern Canada (Lovrich et al., 1997). In laboratory studies, crab smaller than 55 mm carapace width were at high risk of being cannibalized when housed with larger crab (Lovrich et al., 1997). Crab larger than 55 mm carapace width were much less likely to be cannibalized, but the frequency of injury could be high. Here we developed an index of cannibalism based on two aspects of the spatial distribution of snow crab: the overlap of crab smaller than 55 mm carapace width with crab larger than 95 mm carapace width (Figure 28) and the density of crab larger than 95 mm carapace width within the shared space. The proportion of 55 mm carapace width crab in the overlapping area represents the 'exposure' of the smaller population to cannibalism and the density of crab larger than 95 mm carapace width within that area represents the potential 'intensity' of cannibalism in the shared area. We calculated an index of cannibalism over time as the product of exposure and intensity. Consequently, a scenario in which there was large overlap, but low densities of large crab would result in a low cannibalism index value. Similarly, a scenario in which there was low overlap, but high densities would result in a low cannibalism index value. This produces an index that is comparable with estimated mortality–a higher cannibalism index would be expected to be associated with higher mortality if cannibalism is a strong driver of mortality in the size ranges of crabs modeled here.

The proportion of 55 mm carapace width crab overlapping with larger than 95 mm carapace width crab was calculated by finding the intersection of the station IDs at which at least one crab of both size classes was observed. The density of crab larger than 95 mm carapace width was calculated as the number of >95 mm carapace width crab observed at those stations multiplied by the area swept. This exercise was also done by 5 mm size bins to show the overlap of small crab of different sizes with large crab (Figure 29). The final index aggregated all crab smaller than 55 mm carapace width (Figure 30).

Fisheries data

Snow crab are caught both in a directed fishery (i.e. a fishery aimed at capturing snow crab) and non-directed fisheries (i.e. fisheries with targets other than snow crab). In the directed fishery, under-sized and/or dirty shelled crab are often discarded. Snow crab are discarded from non-directed fisheries using a variety of gear types (including trawl, pots, hook-and-line) and targeting a variety of species (e.g. Pacific cod, walleye pollock, and yellowfin sole) that operate over a wide fraction of the Bering Sea shelf (Figure 31). Figure 31 is plotted in log space, so it appears that the bycatch is spread widely over the shelf, but in normal space, the bycatch is more concentrated (e.g. Figure 32). The location of the centroids of the bycatch have moved over time and increases in latitude correspond with warm years in which reduced ice extent allowed for fishing farther north (Figure 33). Bycatch in groundfish trawl fisheries are by far the largest sources of bycatch mortality (Figure 34). Data on discards and bycatch of snow crab are collected by at-sea observers on fishing boats and the percent observer coverage ranges from 10% to 100%, depending on the fishery. Indices of the relative mortality imposed by fisheries discards and bycatch were calculated here as the ratio of the observed numbers of crab discarded or bycaught in a given year divided by the estimated population numbers in a given year. Only discard mortality is considered for the directed fishery in our models because the range of sizes modeled exclude the largest males, which are the targets of the commercial fishery for snow crab.

Generalized additive models

Generalized additive models (GAMs) were used in the R programming language (package mgcv; Wood, 2011) to relate changes estimated mortality by maturity state and year, $m_{p,y}$ to environmental covariates by maturity state, $\phi_{m,y}$, because of their flexibility in fitting potential non-linear relationships. Models were first fitted in which all relevant covariates were included in the model of the form:

$$m_{p,y} = s(\phi_{m,y}) + \epsilon_i \tag{18}$$

where 's()' is a smoothing function based on thin-plate splines, ϕ is a matrix of environmental covariates scaled to mean 0 and standard deviation 1, and ϵ is normally distributed error. The number of knots allowed in the thin-plate splines were restricted to 3 given the relatively short time series and number of potential stressors. Significance of covariates for the full models can be seen in Table 1 and Table 2 and the resulting smooths in Figure 35 and Figure 36. Model diagnostics were acceptable given relatively short time series (Figure 37 & Figure 38). Leave-one out cross validation was performed for the models by systematically excluding a year of data, refitting the model, and recording the deviance explained and significance of the covariates. The consistent significance of specific covariates in this exercise lends some credence that those covariates' influence in the model was not the result of outliers (Figure 2e). Some collinearity existed among covariates (Figure 39 & Figure 40), but none of the collinear variables were significant in the models.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	0.6052	0.0505	11.9819	< 0.0001
B. smooth terms	edf	Ref.df	F-value	p-value
s(temperature)	1.9964	2.3824	4.5850	0.0204
s(disease)	1.0000	1.0000	1.3152	0.2650
s(discard)	1.0000	1.0000	0.4639	0.5036
s(bycatch)	1.0000	1.0000	1.1509	0.2961
s(mat_pop)	1.8356	1.9601	4.0153	0.0261

Table 1: GAM output for full model predicting mature mortality. Deviance explained = 66.8%

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	0.1750	0.0121	14.4737	< 0.0001
B. smooth terms	edf	Ref.df	F-value	p-value
s(disease)	1.6017	1.8194	1.3939	0.3717
s(temperature)	1.5788	1.7957	6.0398	0.0067
s(mat_pop)	1.9039	1.9757	4.5453	0.0352
s(predation)	1.0000	1.0000	0.1465	0.7064
s(bycatch)	1.0000	1.0000	1.1798	0.2917
s(cannibalism)	1.6038	1.8388	1.5158	0.3215

Table 2: GAM output for full model predicting immature mortality. Deviance explained = 72.2%

Models that excluded insignificant variables from each full model were used in out-of-sample prediction and randomization tests (see Table 3 & Table 4 for covariate significance and deviance explained and Figure 41 & Figure 42 for model diagnostics). One thousand iterations of a randomization test were performed in which the covariate time series were randomized, the models refit, and the deviance explained recorded. This test was aimed at understanding if the explanatory power of the model was a result of the number of covariates considered and the flexibility of the model or if the results were an indication of some underlying signal in the data. If the deviance explained by the model using the non-randomized data exceeded the 95th quantile of the randomization trials, the deviance explained from the fitted model is less likely to be a result of over-fitting resulting from too many covariates or too flexible smooths. The deviance explained from both of the trimmed models exceeded the 95th quantile of deviance explained from the randomization (Figure 43 & Figure 44). Out-of-sample predictions were made by excluding the last 1,2, and 3 years of data, refitting

the model, then attempting to predict the held out data based on the covariates observed in those years (see figure 2 of the main text for a discussion).

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	0.6052	0.0497	12.1798	< 0.0001
B. smooth terms	edf	Ref.df	F-value	p-value
s(temperature)	1.8723	2.2591	4.5071	0.0212
s(mat_pop)	1.8819	1.9750	7.4877	0.0025

Table 3: GAM output for trimmed model predicting mature mortality. Deviance explained = 62.9%

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	0.1750	0.0130	13.4588	< 0.0001
B. smooth terms	edf	Ref.df	F-value	p-value
s(temperature)	1.6929	1.9043	8.0575	0.0016
s(mat_pop)	1.9760	1.9981	6.2179	0.0077

Table 4: GAM output for full model predicting immature mortality. Deviance explained = 59.1%

How could temperature relate to mortality mechanistically?

Increased temperature was consistently correlated with increased estimated mortality in our models, but the range of temperatures observed were not beyond the thermal tolerances of snow crab. Foyle et al. (1989) captured 20 snow crab of carapace size 85-95 mm in 1986 and raised them in the lab in a range of thermal regimes to understand the impacts of increased temperatures on mortality and caloric requirements for snow crab. In addition to identifying the thermal tolerances of snow crab (crab stop eating around 12 degrees C), Foyle et al. observed a doubling of caloric requirements for snow crab held in 3 degrees Celsius water as compared to those in 0 degree waters. Here we calculated an index of the caloric requirements for the population of snow crab in the eastern Bering Sea over time using the abundance at size of snow crab observed in the NMFS survey, the temperature occupied of crab at size calculated from observations of bottom temperature in the NFMS survey, and the observations of caloric requirements of snow crab by temperature produced by Foyle et al. (1989). The relationship between temperature and the caloric requirements of snow crab ($kCal_t$) reported by Foyle et al. was:

$$kCal_{s=90mm,t} = 2.2 * e^{\frac{-(t-5.2)^2}{30.7}} \tag{19}$$

Snow crab numbers at size (s) by year (y) ($N_{s,y}$) and the temperature occupied at size by year ($T_{s,y}$ were calculated as described above. The caloric requirements reported in Foyle et al. were based on observations of crab that were 85-95 mm carapace width, so these results need to be extrapolated to the range of sizes used in this analysis. Kleiber's law (Kleiber, 1947) states there is a consistent relationship between the body mass and metabolic requirements of organisms (kCal). The relationship has been generalized as:

$$kCal_m = mass^{0.75} \tag{20}$$

Calculating the metabolic requirements for snow crab at size by year, $kCal^{snow}_{s,y}$, can be calculated by evaluating the caloric requirements of 90mm carapace width crab at a given temperature were calculated, then scaling that up or down based on Kleiber's law:

$$kCal^{snow}_{s,y} = \frac{2.2 * e^{\frac{-(t-5.2)^2}{30.7}}}{300^{0.75}} w_s^{0.75} \tag{21}$$

The resulting caloric requirements by size and temperature can be seen in Figure 3b of the main text. The population-wide caloric requirements increased sharply in 2018 and to explore potential impacts of this increase, we analyzed the weight at size data available (Figure 45). GAMs were used to predict observed weights at size $w_{i,s,y}$ using the bottom temperature in which the crab was collected, t_i, measured carapace width cw_i, and year as a factor:

$$w_{i,s,y} = s(cw_i) + s(t_i) + year + \epsilon \tag{22}$$

The GAMs explained 97.4% of the deviance in the weights of snow crab and all covariates were significant (Table 5).

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	218.5199	2.2252	98.2019	< 0.0001
as.factor(AKFIN_SURVEY_YEAR)2015	6.4525	3.1690	2.0361	0.0419
as.factor(AKFIN_SURVEY_YEAR)2017	12.6093	2.4840	5.0763	< 0.0001
as.factor(AKFIN_SURVEY_YEAR)2018	-11.9217	6.2536	-1.9064	0.0568
as.factor(AKFIN_SURVEY_YEAR)2019	4.0886	2.7473	1.4882	0.1369
B. smooth terms	edf	Ref.df	F-value	p-value
s(WIDTH)	6.4225	7.5862	6340.9617	< 0.0001
s(GEAR_TEMPERATURE)	1.9362	2.3359	17.0800	< 0.0001

Table 5: GAM output for model predicting male snow crab weight. Deviance explained = 97.4%

In general, higher temperatures were associated with higher weight at size (Figure 46). The weight at size curves for 2015 and 2017 were scaled significantly higher than the base year of 2011, whereas the year 2018 was marginally significantly lower (p=0.057). The marginal significance likely resulted from the relatively small sample size of weight at size available in 2018 (N=27), but the effect size was large (the coefficient associated with 2017 was 12.60; the coefficient associated with 2018 was -11.92) which translated to large differences in estimated weight at size between the years reported in the main document.

References:

Alaska Fisheries Science Center, 2022: AFSC/REFM: North Pacific Groundfish Diet Data 1981-present, Aydin, K., https://www.fisheries.noaa.gov/inport/item/20485.

Mueter, F.J., Litzow, M.A. 2008. Sea ice retreat alters the biogeography of the Bering Sea continental shelf. Ecological Applications. 18(2): 309-320.

(more to come)

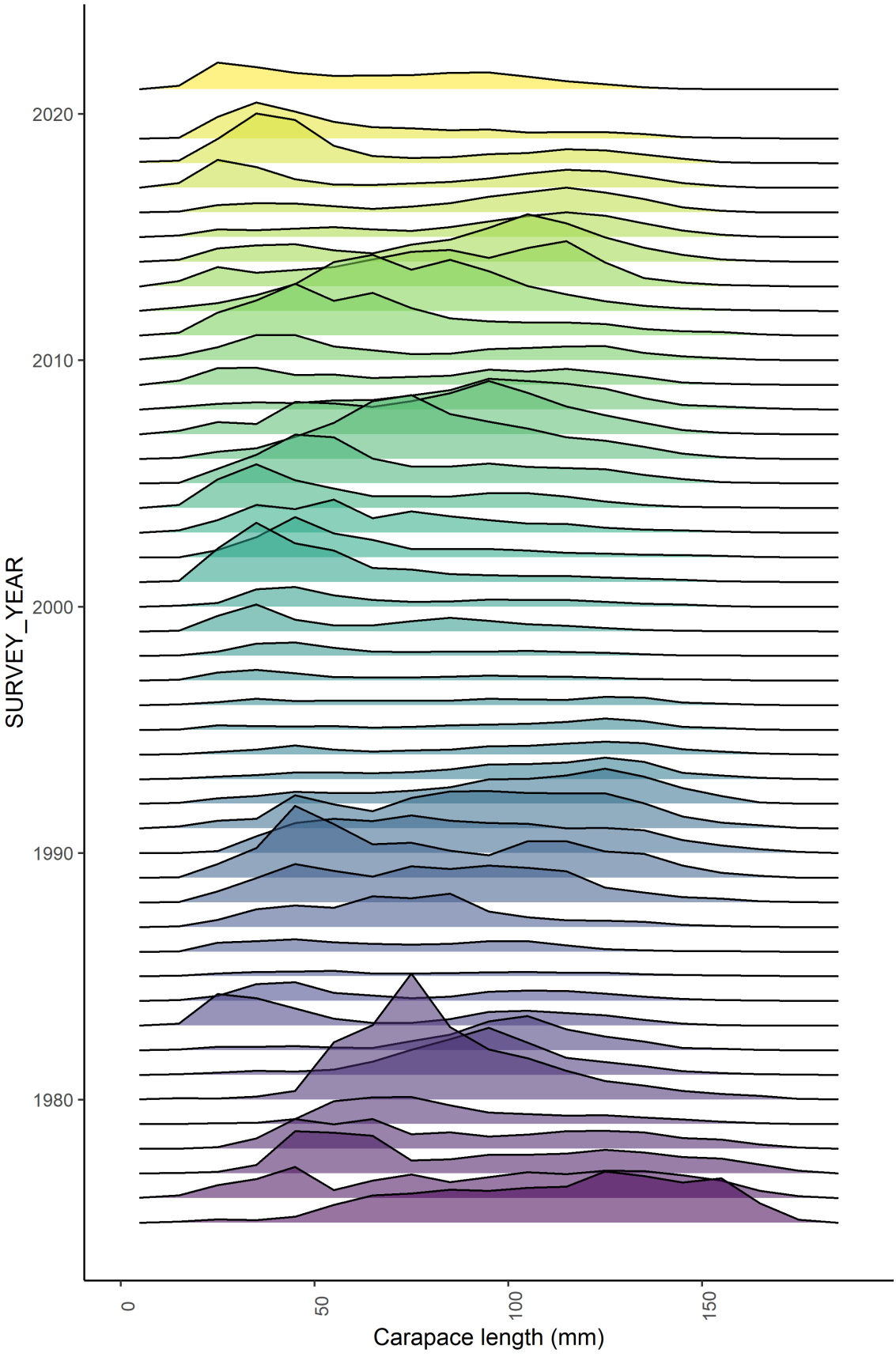

Figure 1: Observed Tanner crab.

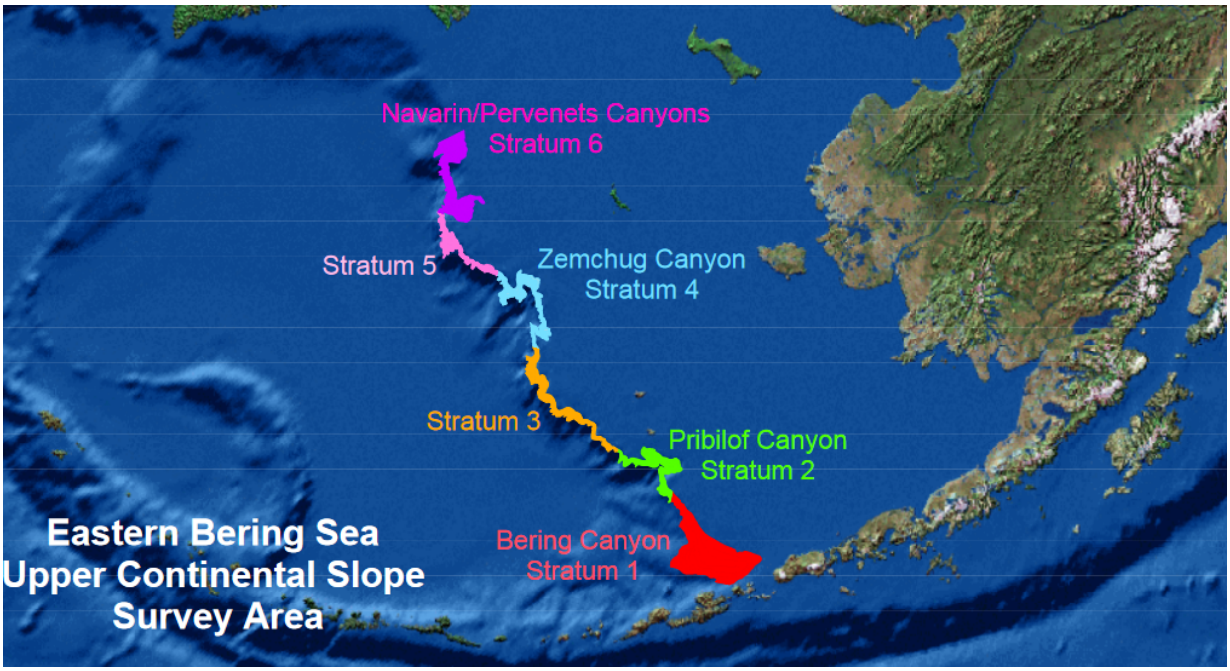

Figure 2: Map of slope habitat.

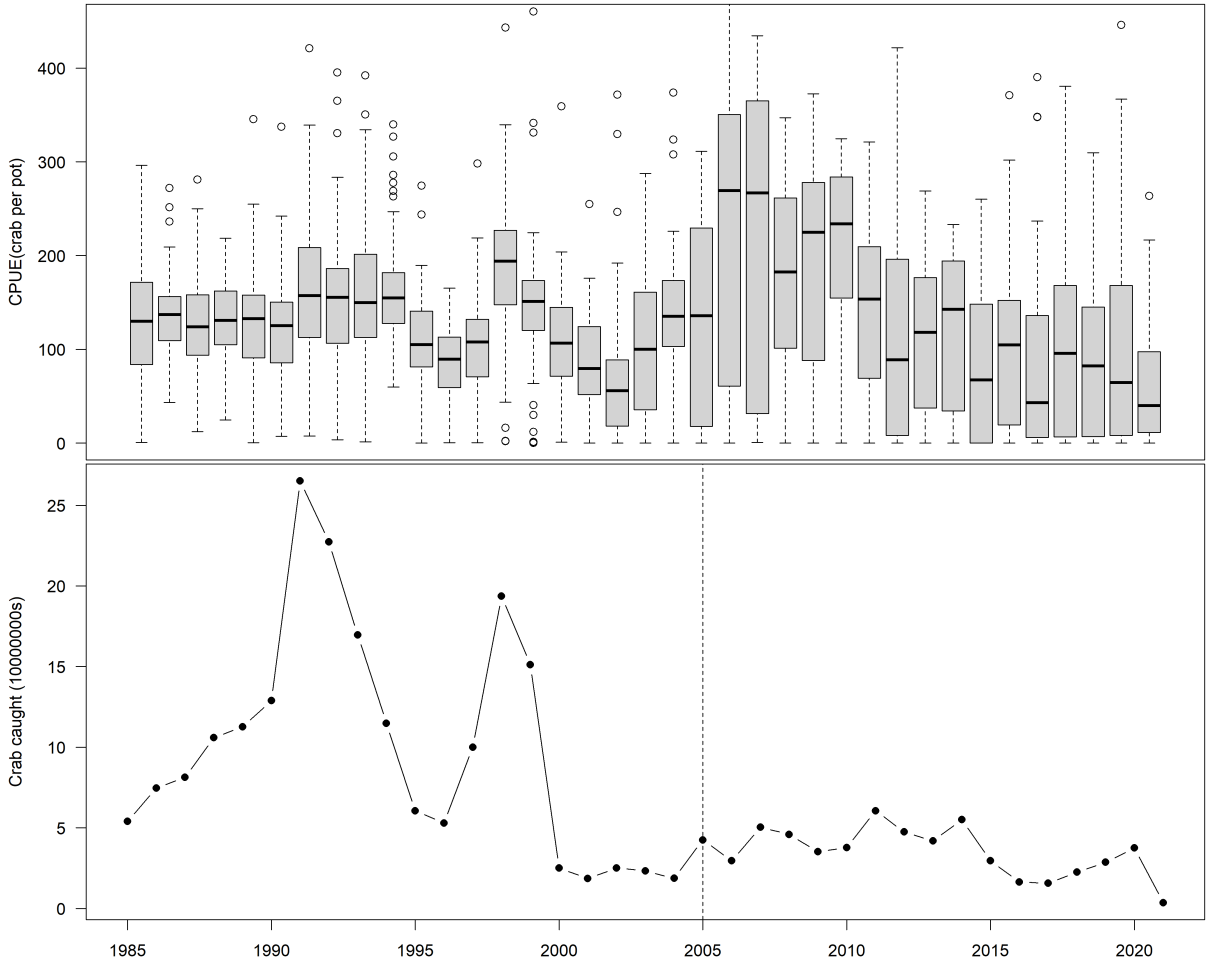

Figure 3: Fishery cpue.

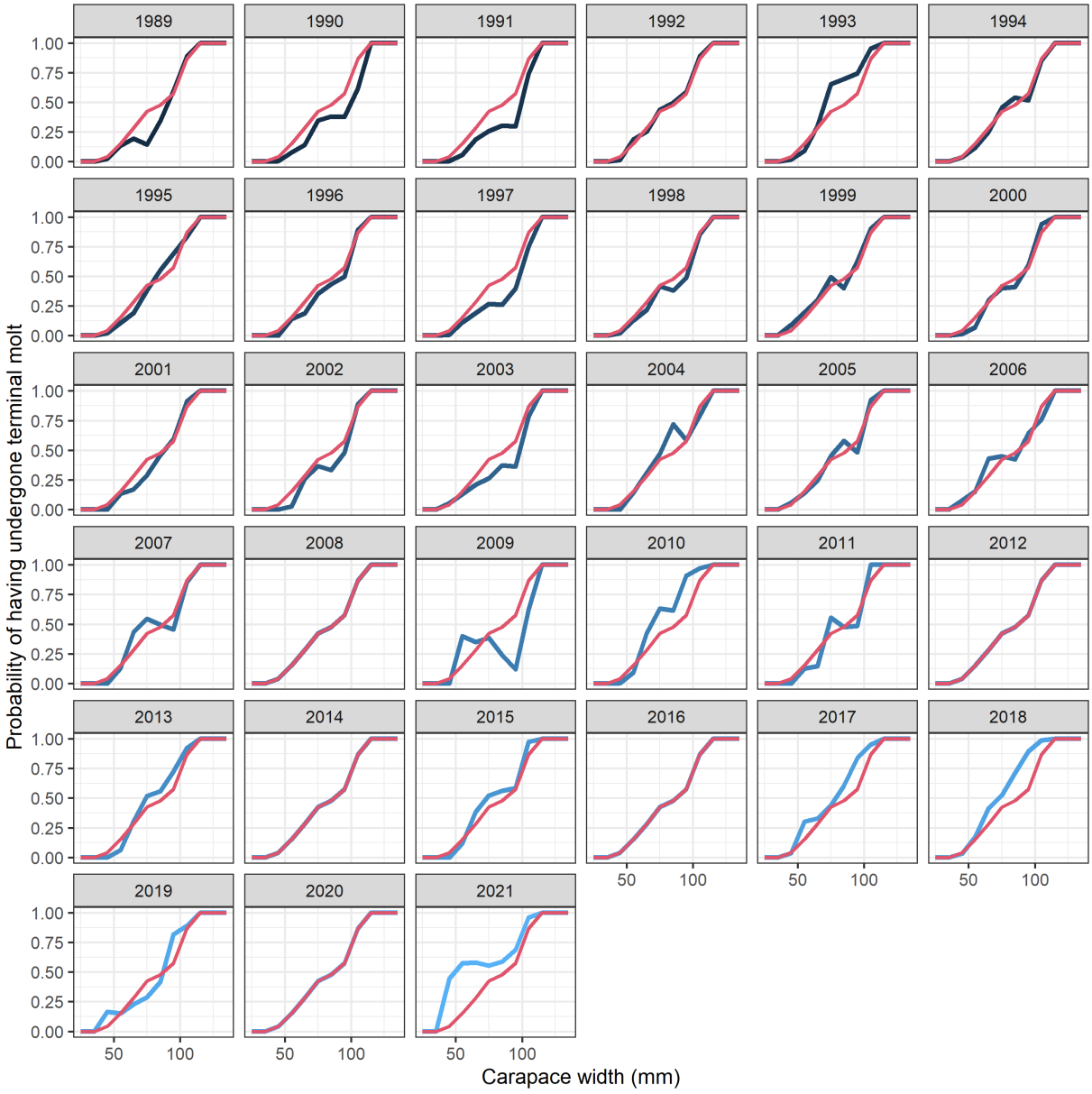

Figure 4: Observed proportion of mature new shell mature crab in the NMFS summer survey. Red line represents the median over years and the blue lines are the observed data. Chela height data were not collected in years without a blue line. These data are used to separate the numbers at size into mature and immature states for the input data to the population dynamics model.

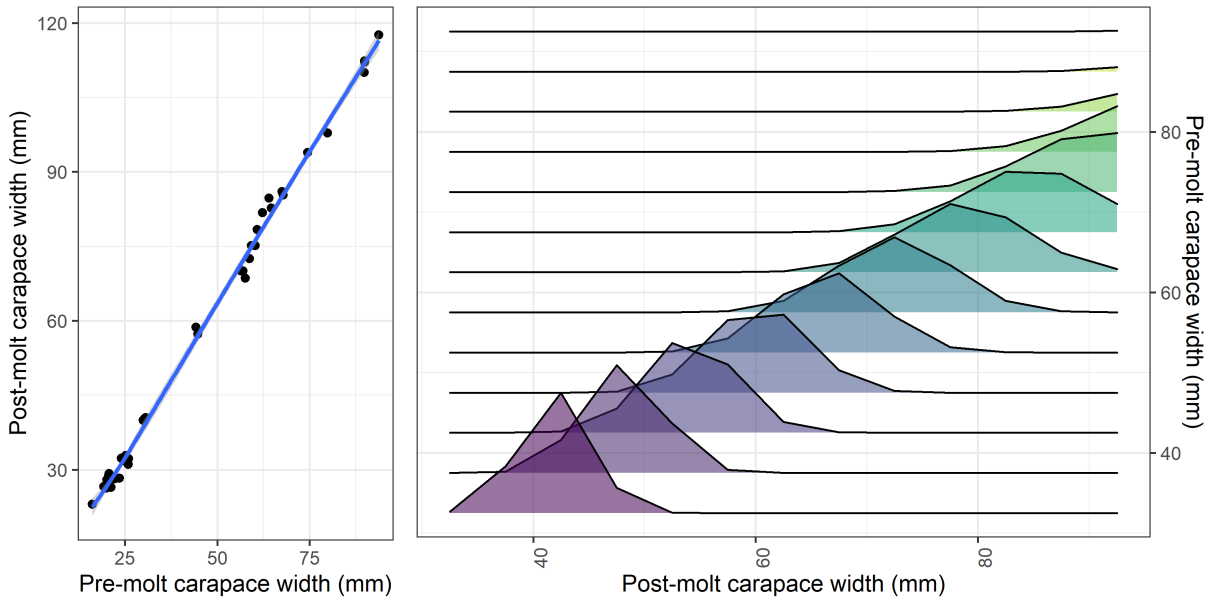

Figure 5: Empirical relationshipe between pre- and post-molt size (left) derived from crab capture in the wild pre-molt and observations in the lab. Calculated size-transition matrix used in the population dynamics model (right).

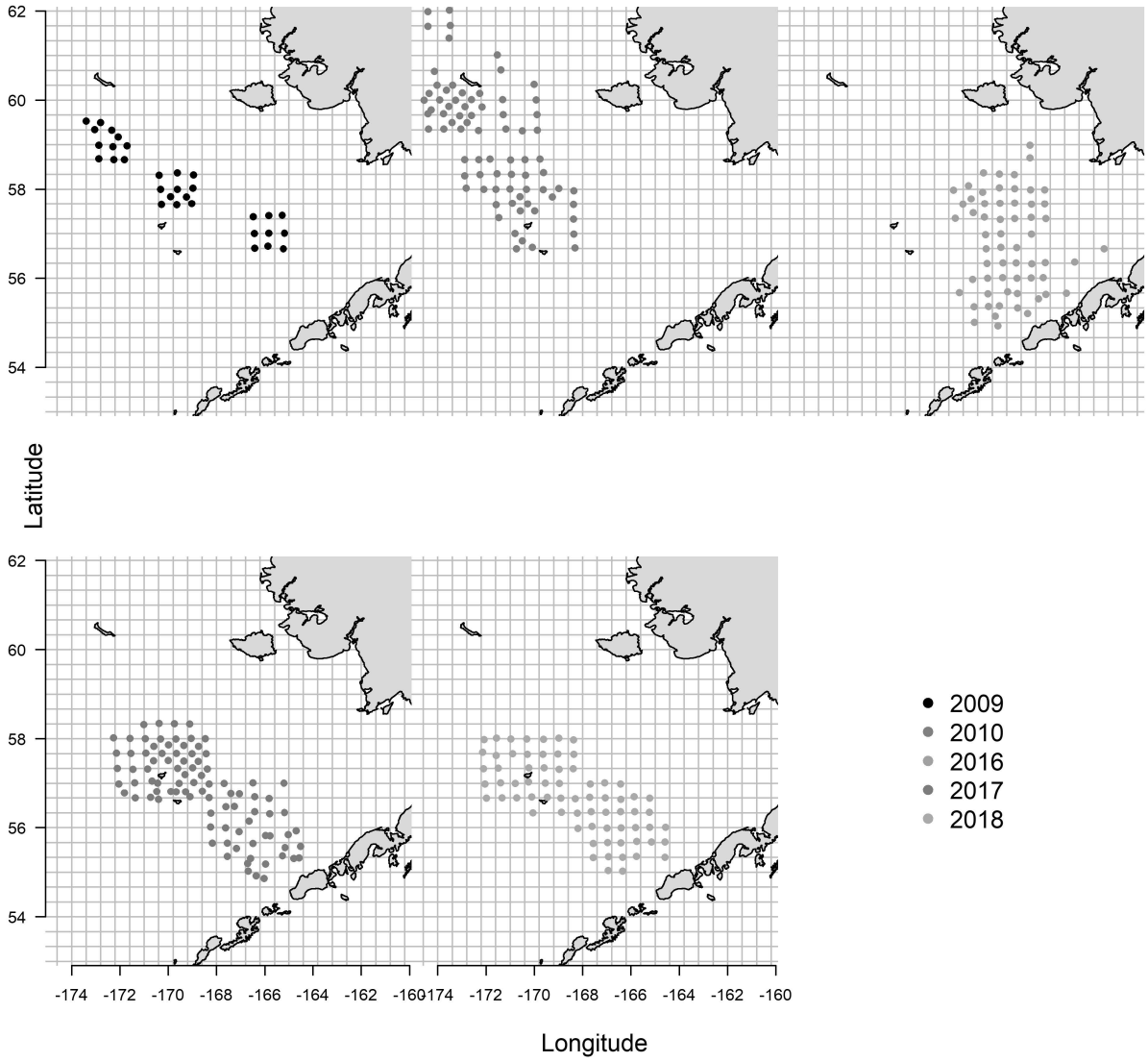

Figure 6: Locations of the BSFRF experimental trawls to evaluate the capture efficiency of the NMFS summer trawl survey for snow crab in the eastern Bering Sea.

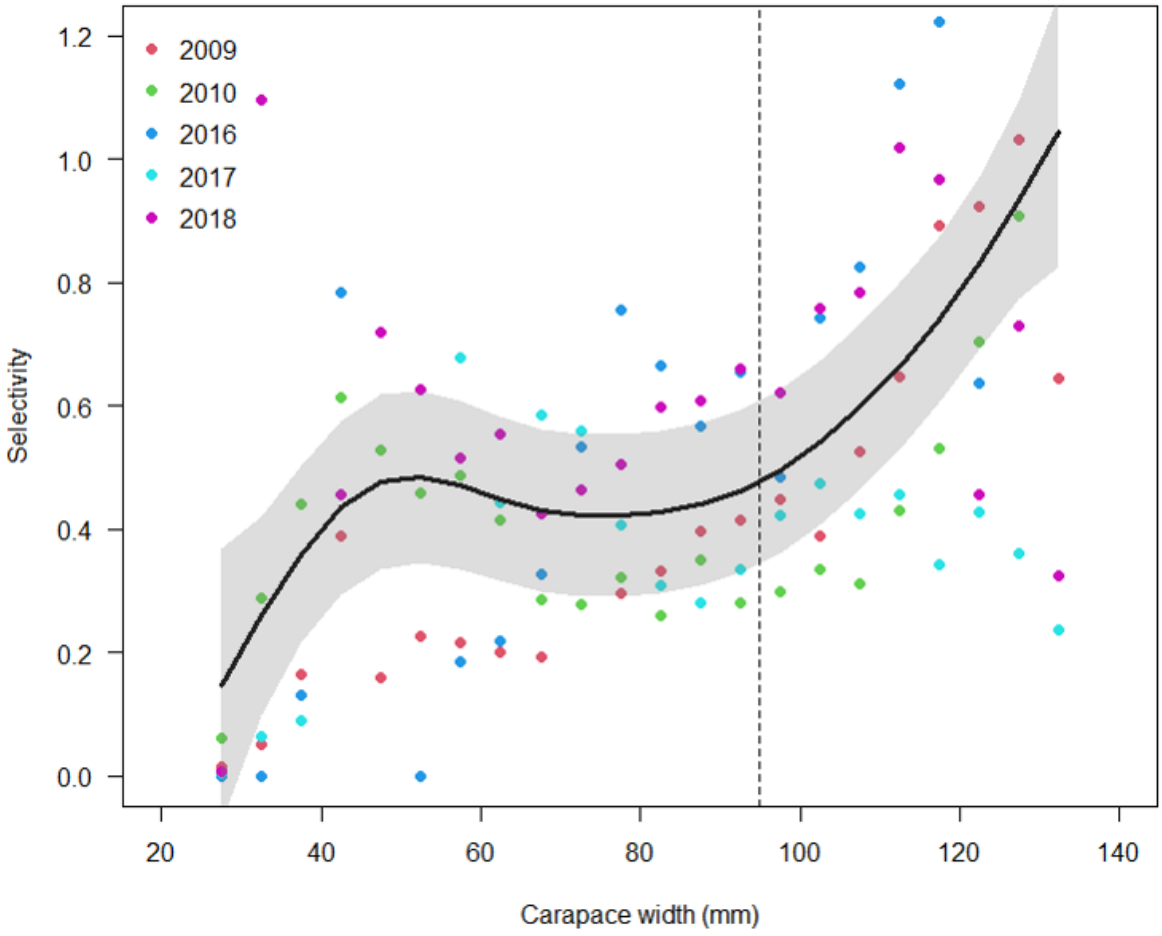

Figure 7: Inferred selectivity from the BSFRF experimental trawls.

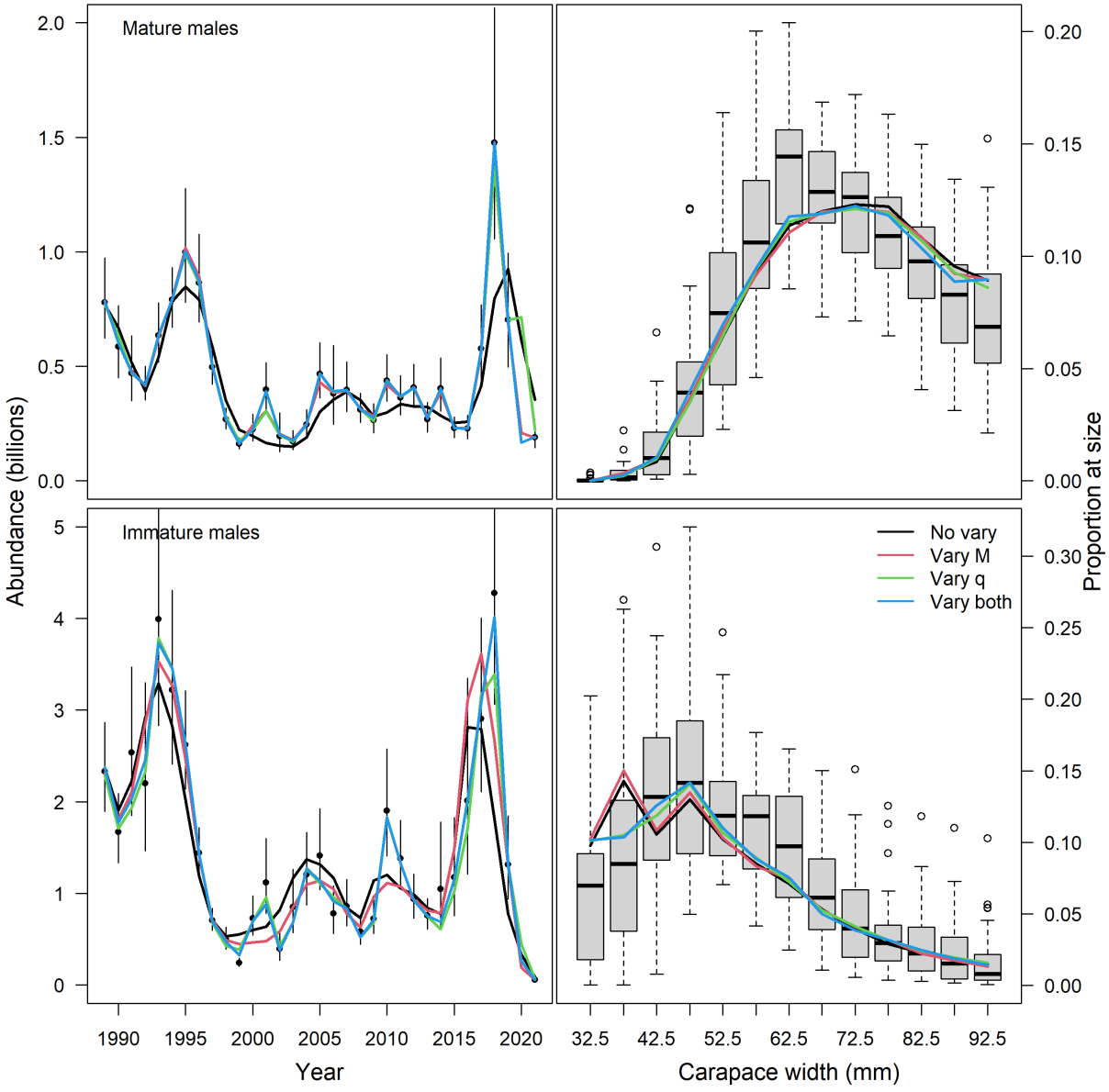

Figure 8: Fits of models with increasing complexity.

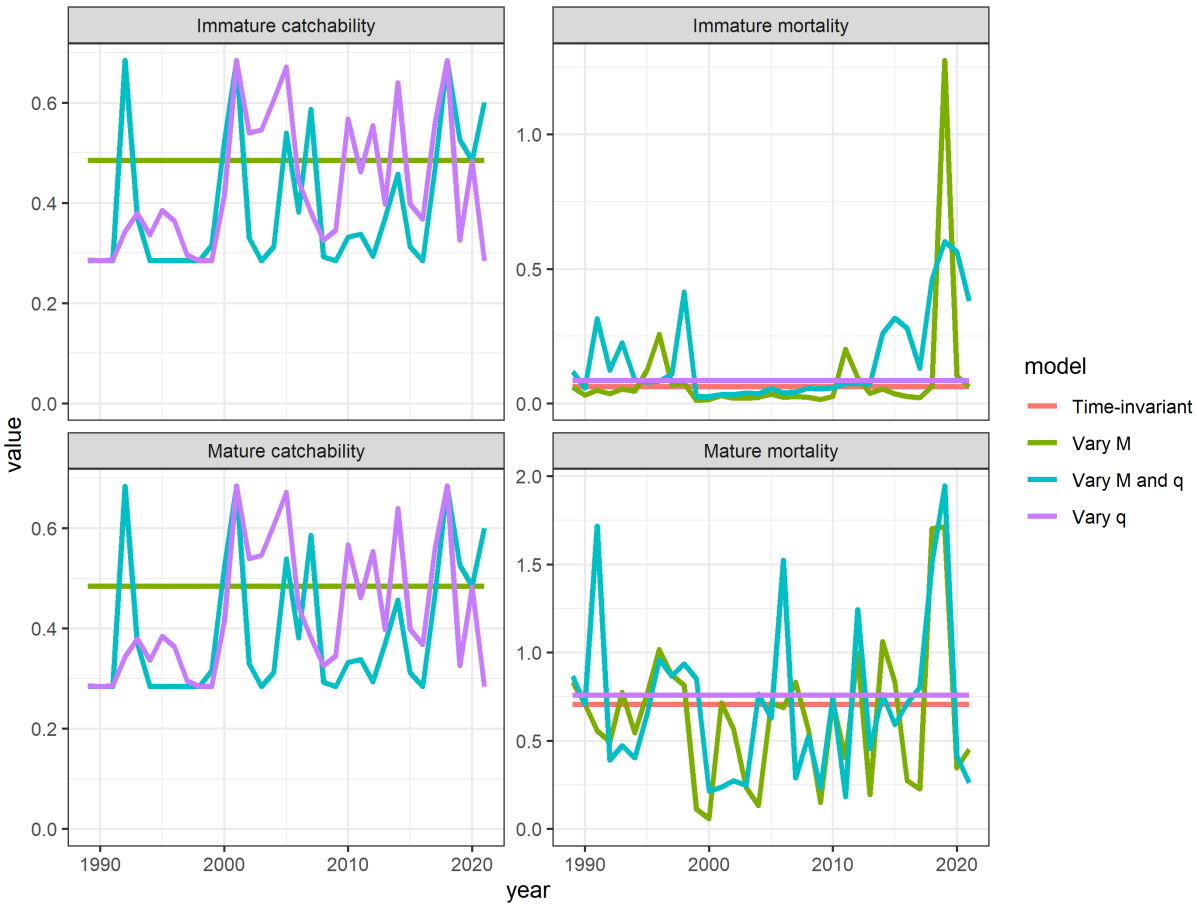

Figure 9: Estimated processes from model with increasingly complex time-variation.

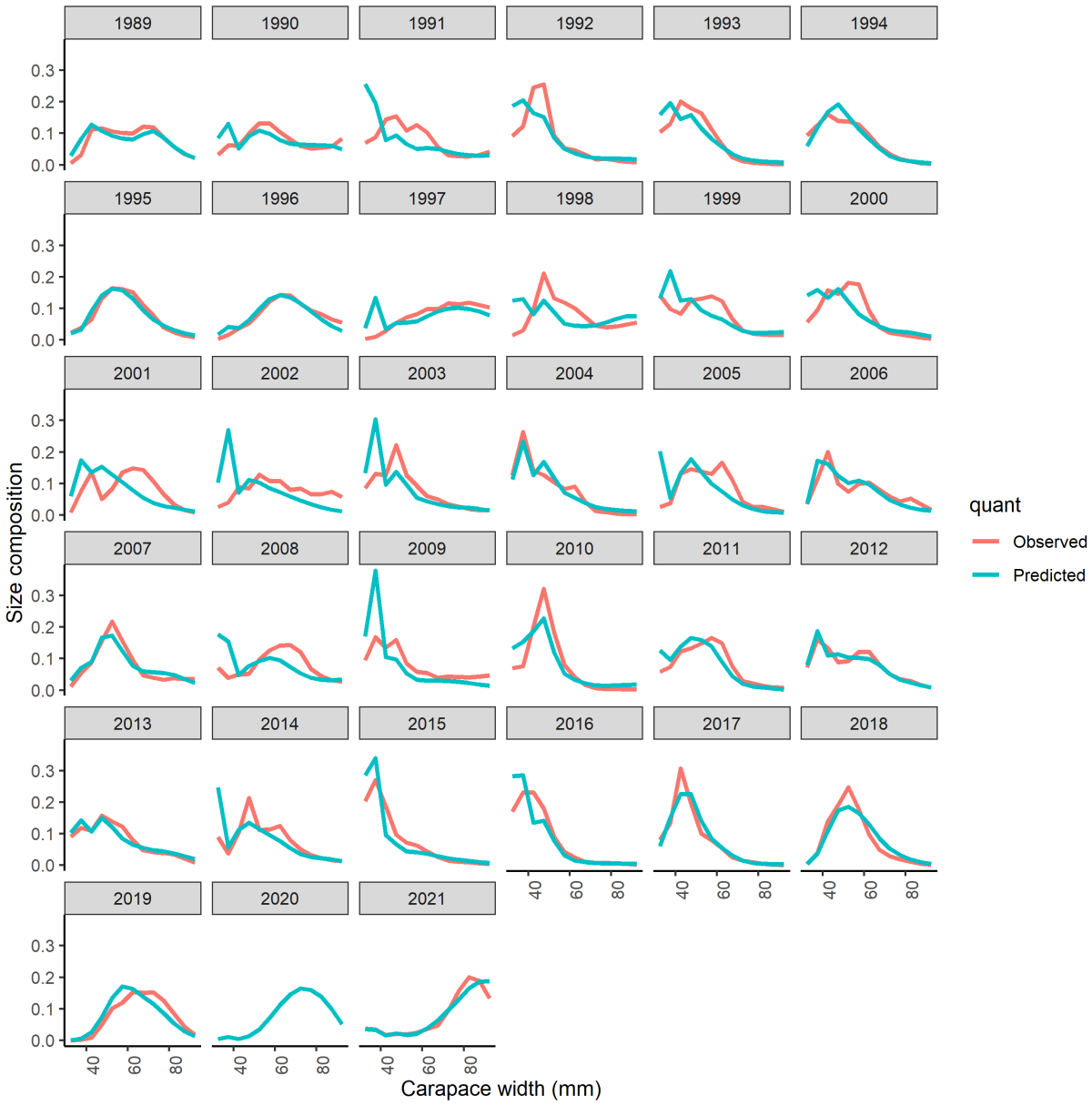

Figure 10: Fits from all years to immature size composition data from a model in which mortality varies over time.

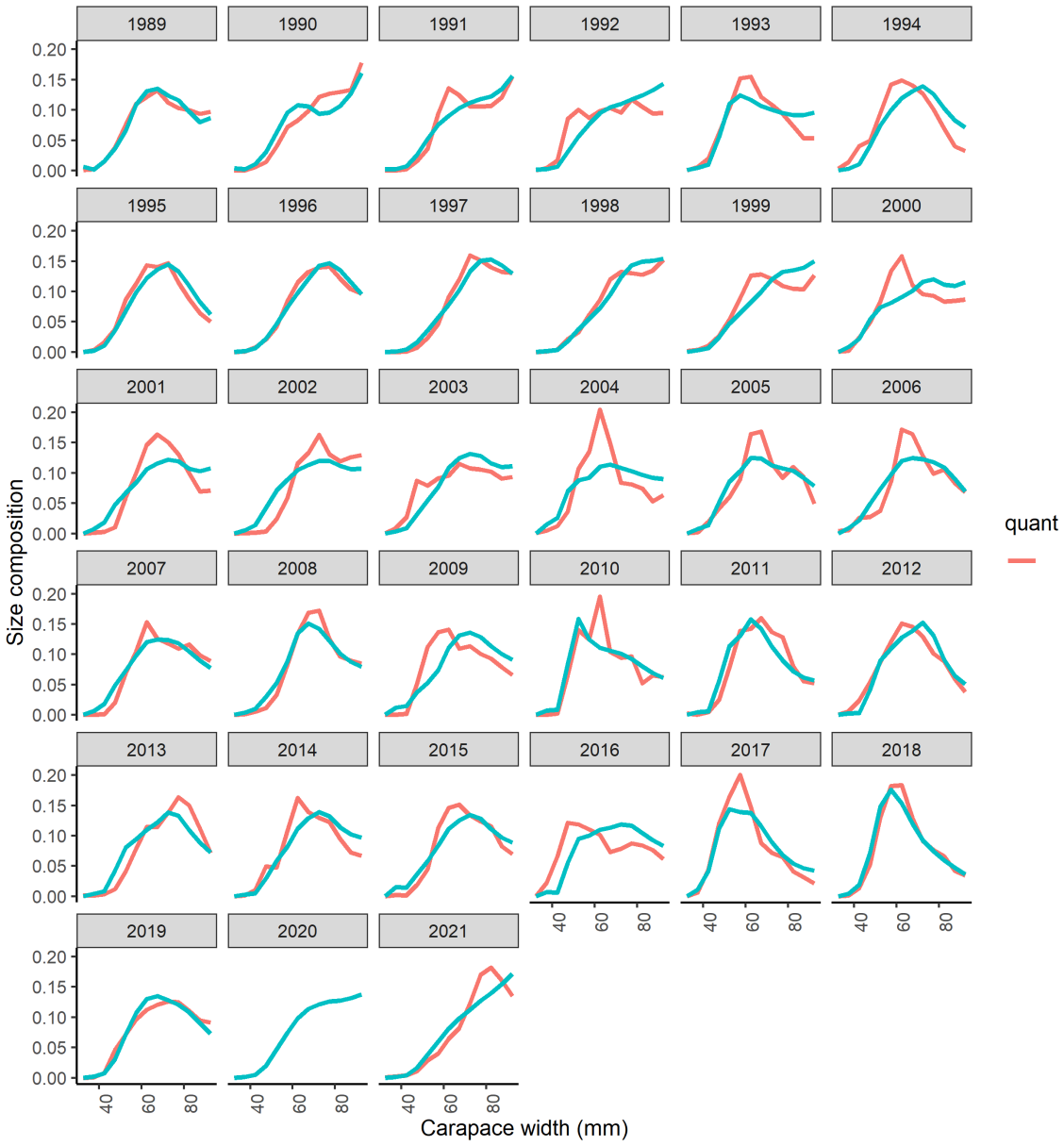

Figure 11: Fits from all years to mature size composition data.

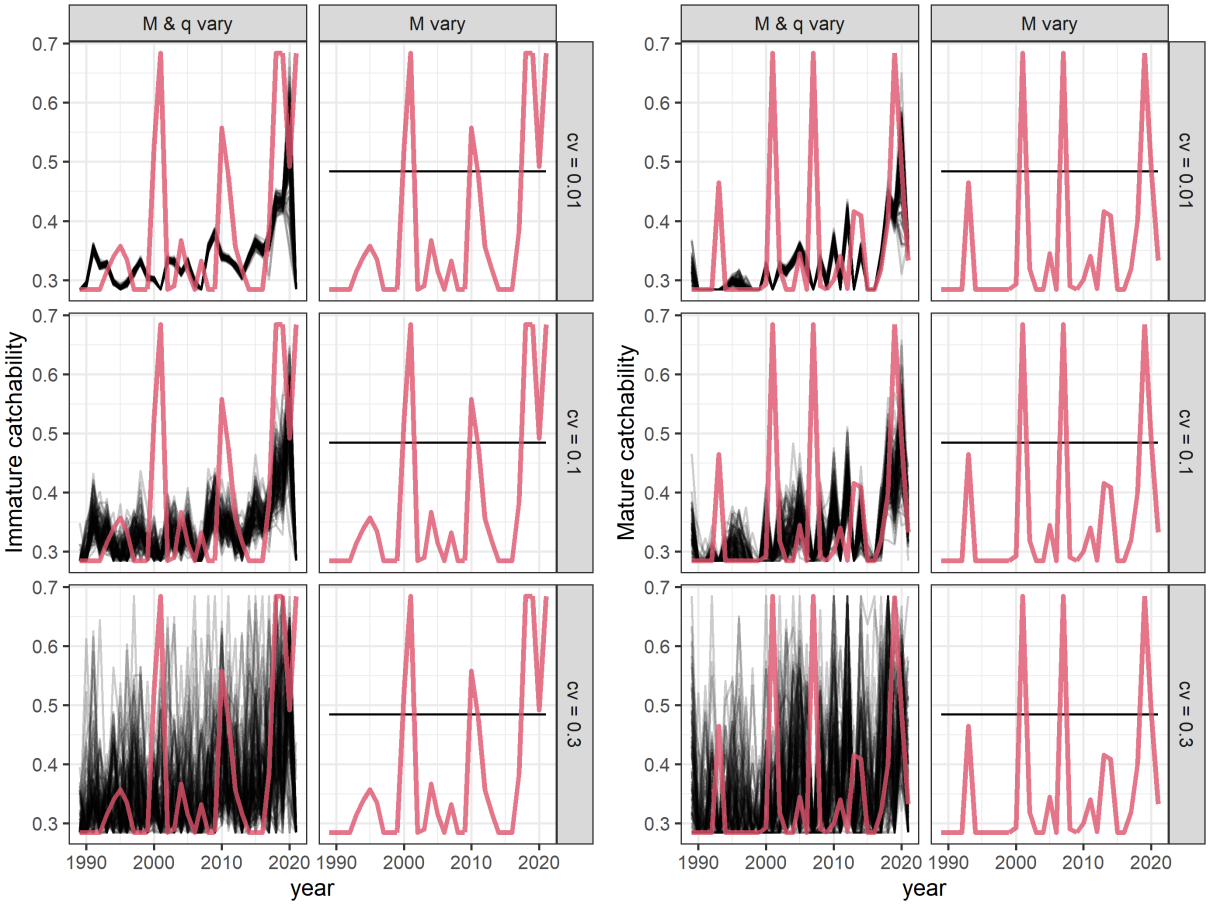

Figure 12: Estimates of cathchability by maturity state (black lines) compared to the underlying values (red line) from simulations testing the estimation ability of the population dynamics models.

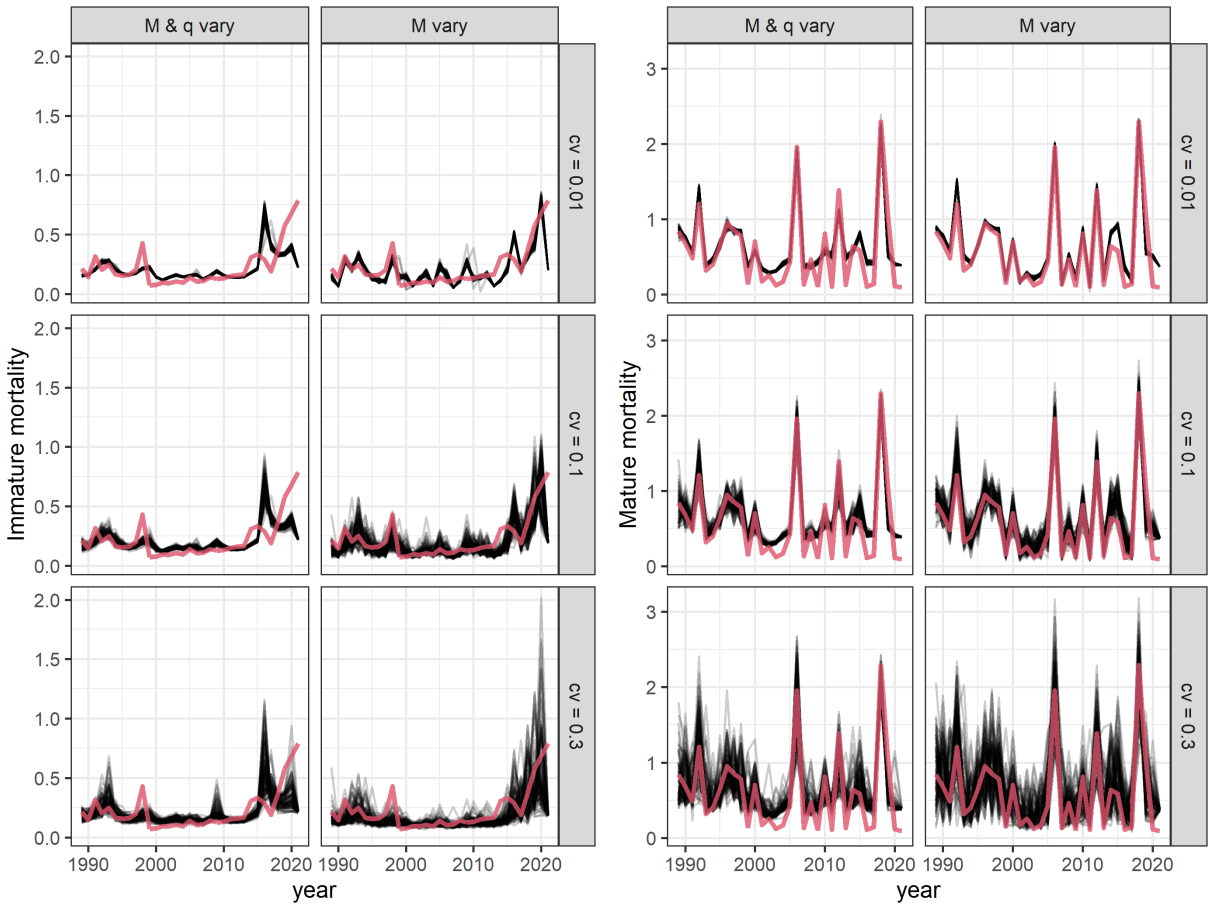

Figure 13: Estimates of mortality by maturity state (black lines) compared to the underlying values (red line) from simulations testing the estimation ability of the population dynamics models.

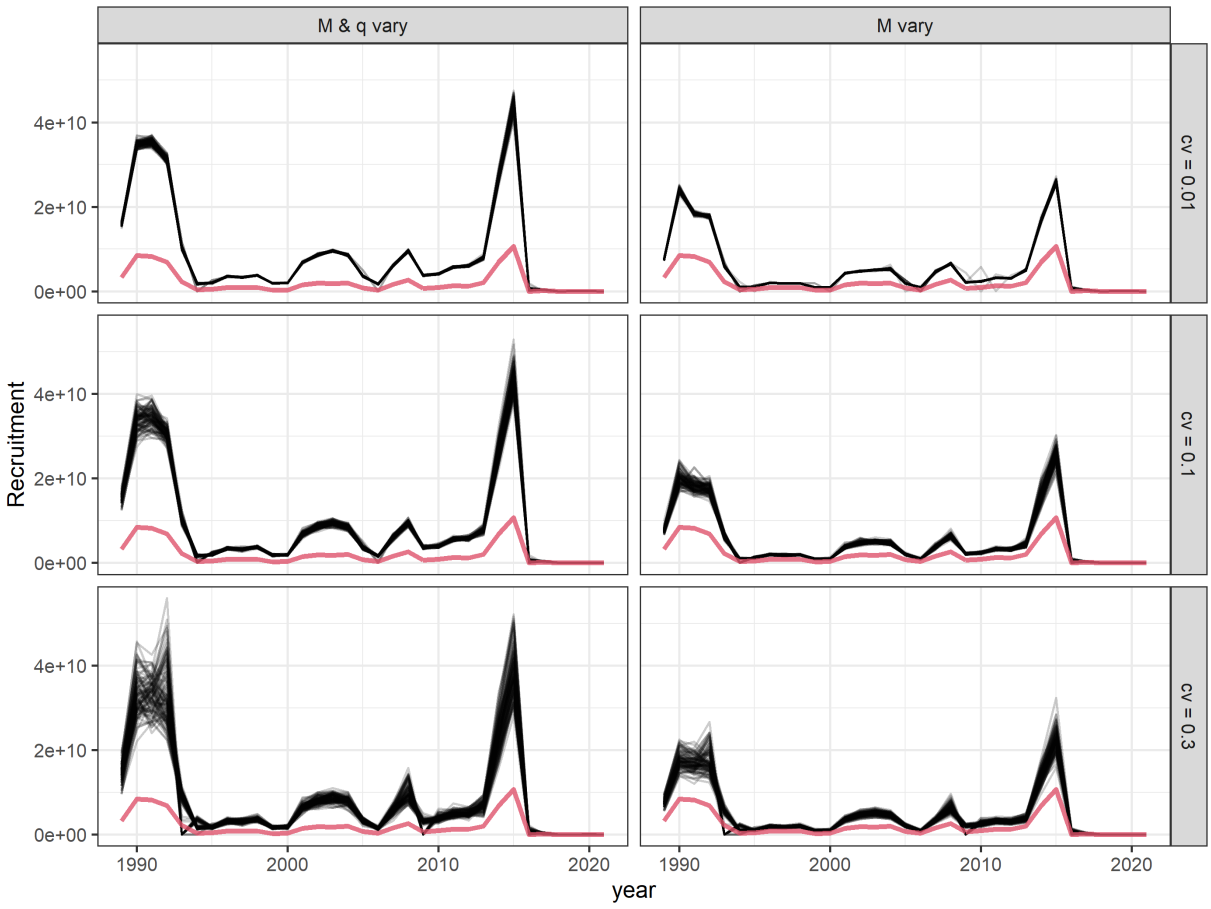

Figure 14: Estimates of recruitment (black lines) compared to the underlying values (red line) from simulations testing the estimation ability of the population dynamics models.

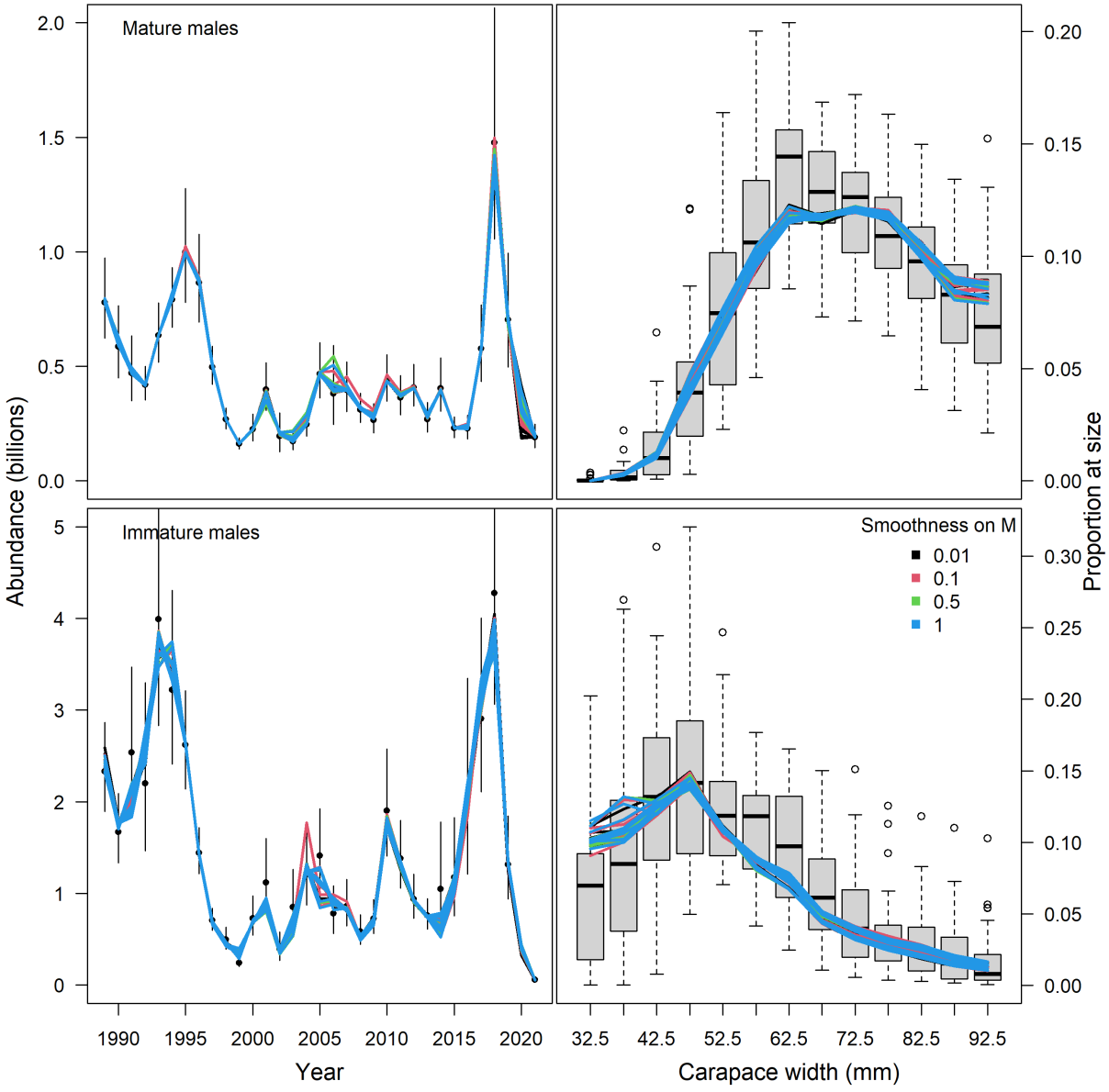

Figure 15: Model fits from sensitivity tests.

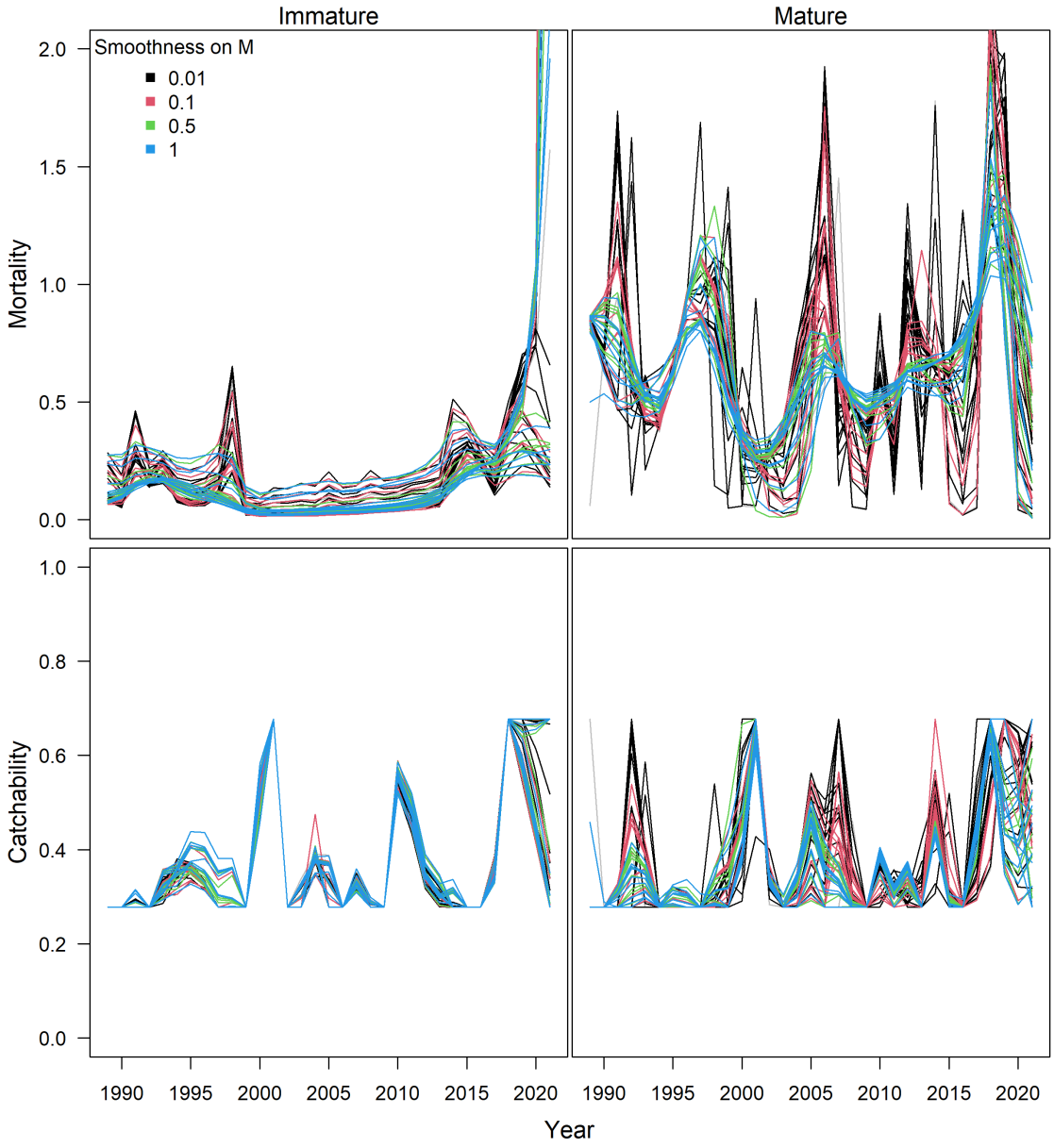

Figure 16: Estimates of mortality and catchability by maturity state over sensitivity runs.

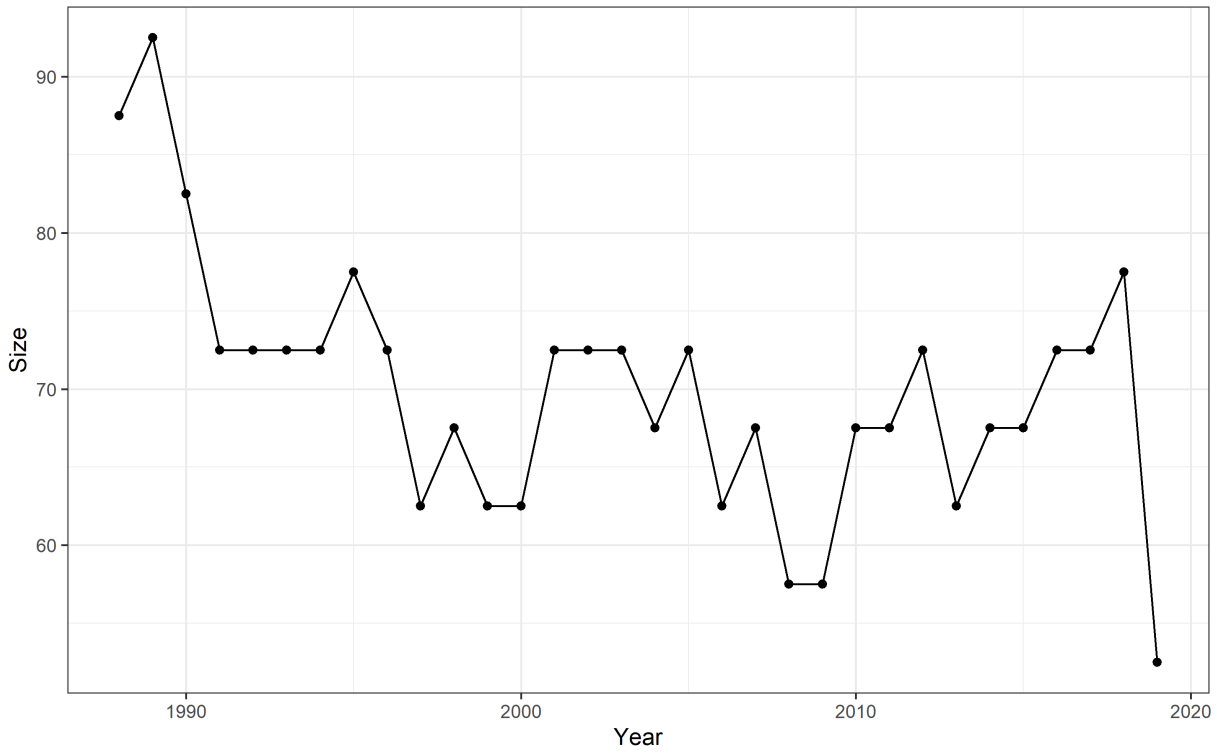

Figure 17: Size at which half of the crab in the population are mature over time. (note, this is not the probability of undergoing terminal molt, rather the proportion of the number of mature vs. immature crab at size in the population)

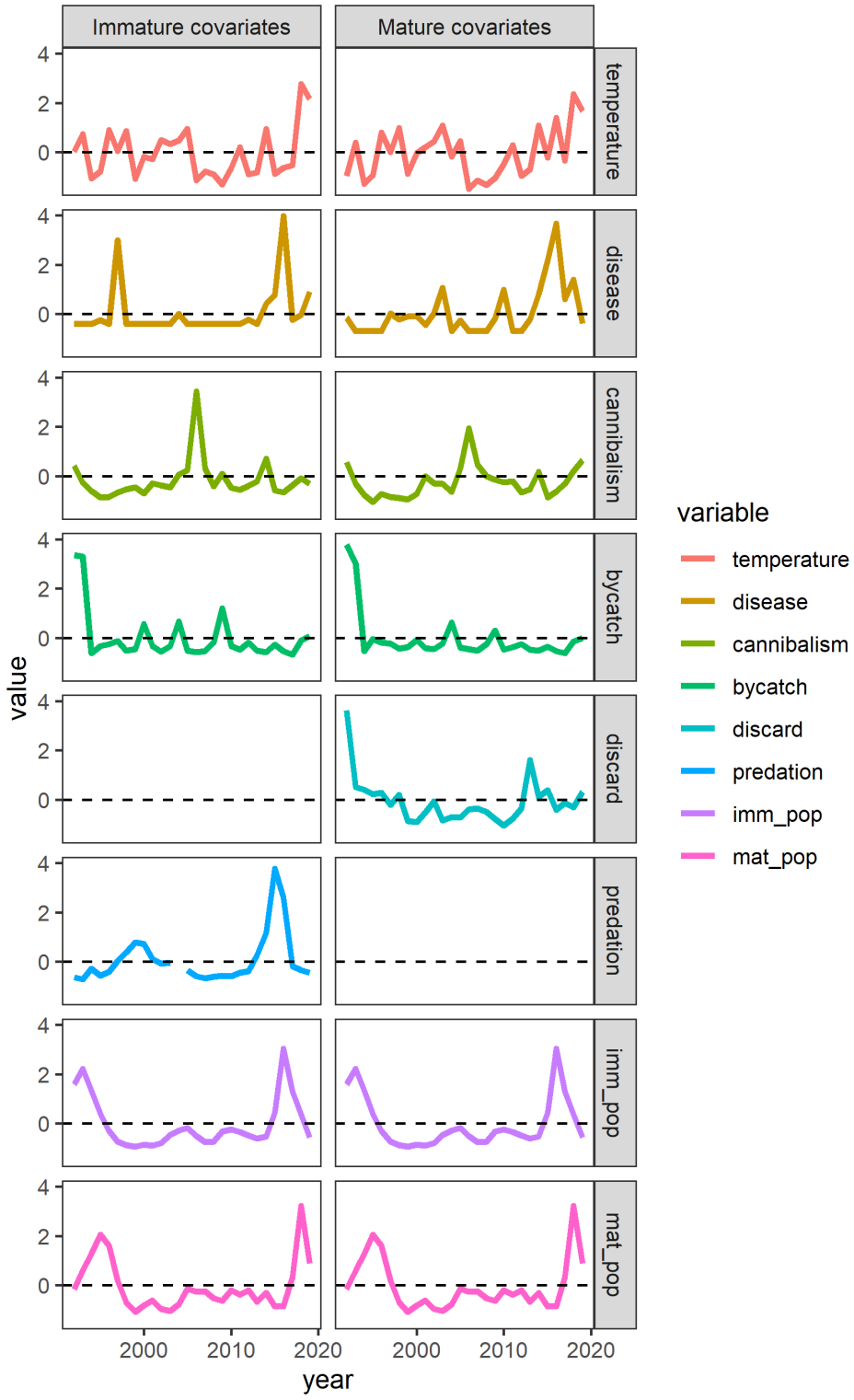

Figure 18: Calculated covariates incorporated into GAMs to relate stressors to estimated mortality. Two covariates (discard and predation) are only relevant for one maturity state based on the critical role size plays in the process (i.e. discards are primarily relatively large crab and predation is primarily smaller crab).

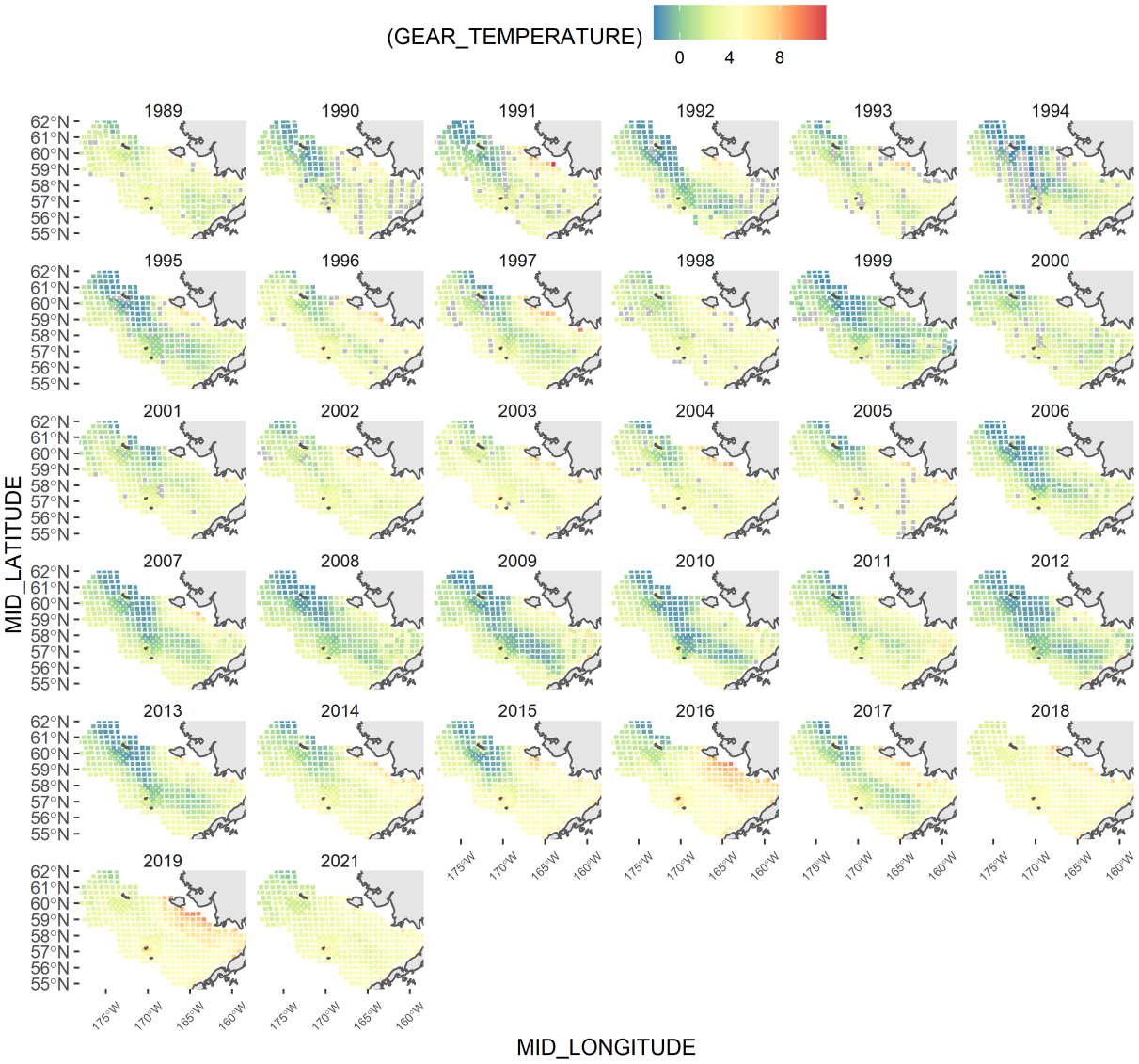

Figure 19: Bottom temperature at the time of the NMFS summer survey.

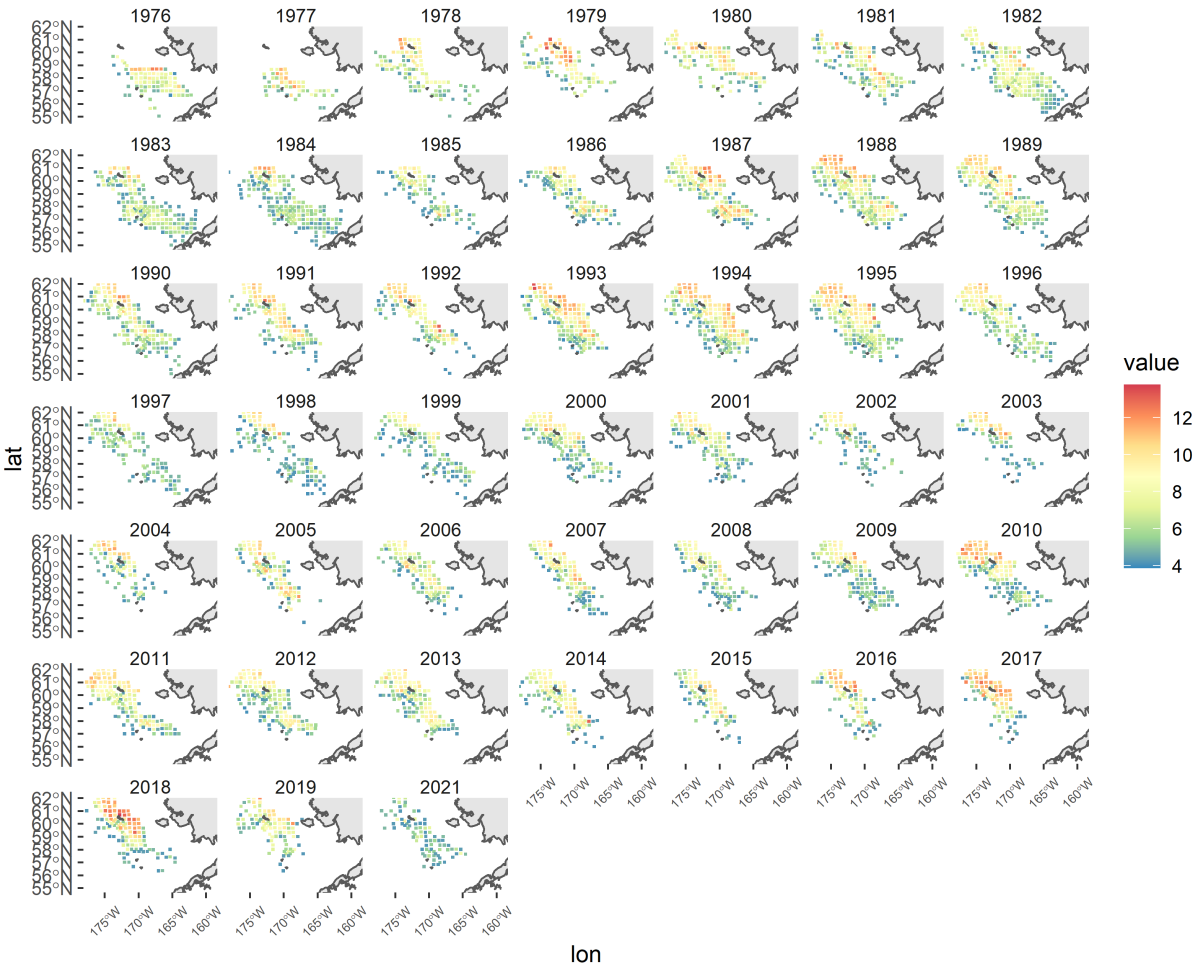

Figure 20: Distribution and intensity of densities of crab <55 mm carapace width in the NMFS summer survey.

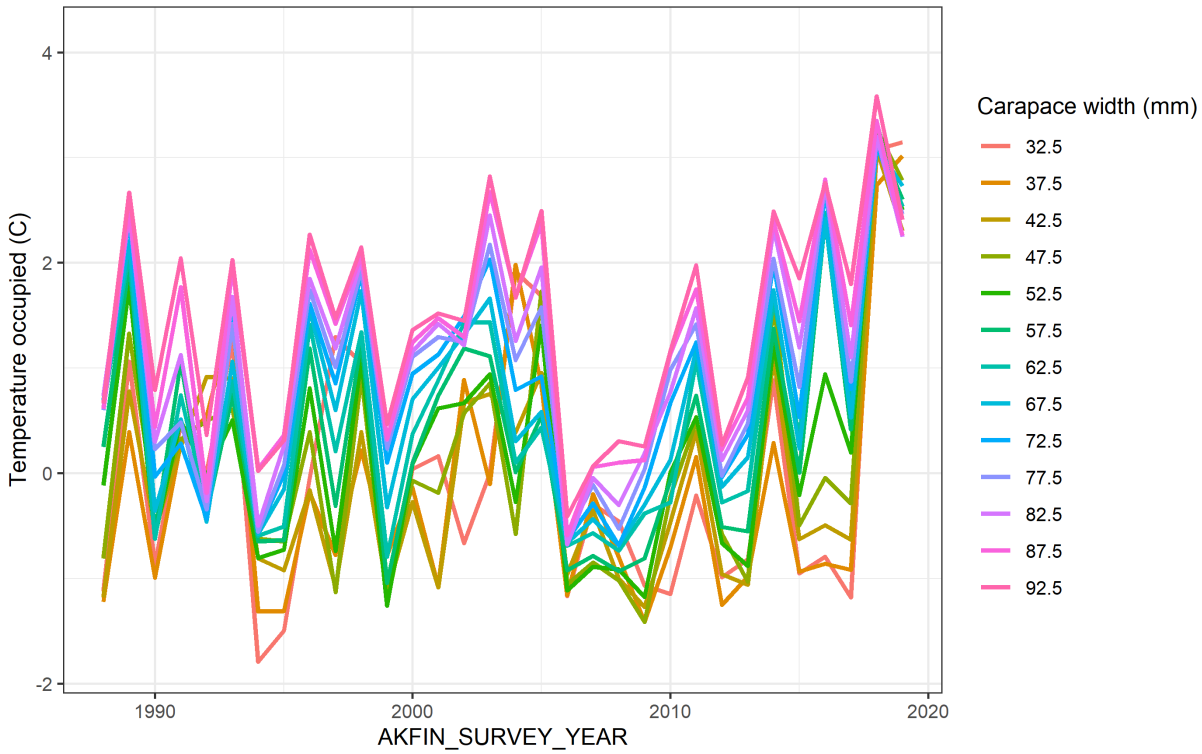

Figure 21: Temperature occupied over time of crab by 5 mm size bin.

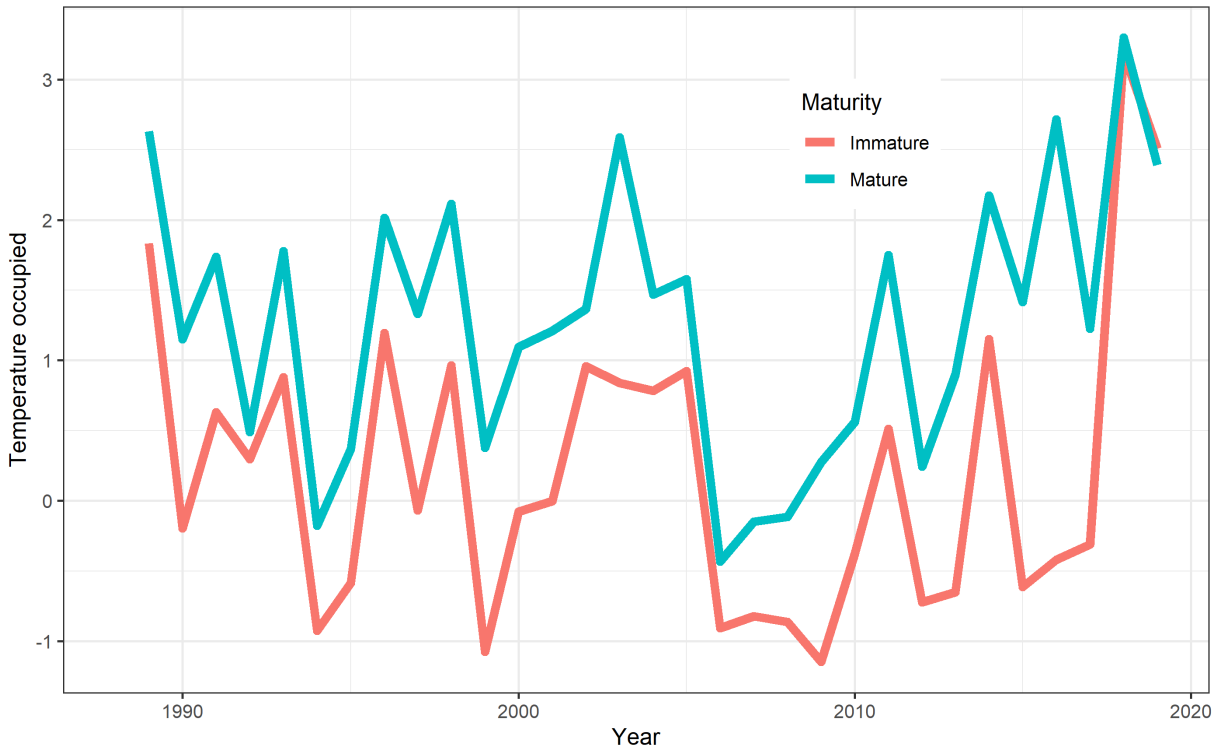

Figure 22: Temperature occupied over time of crab by maturity state.

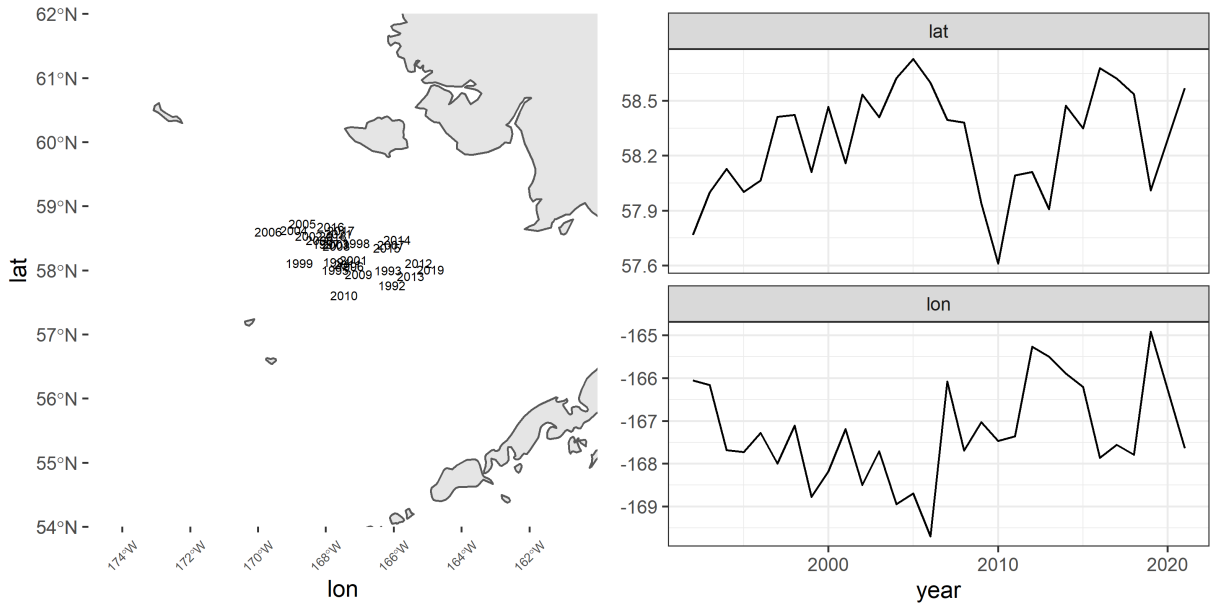

Figure 23: Centroids of abundance for Pacific cod in the Bering Sea over time (left). Right panels show the time series of the centroids broken down by latitudinal and longitudinal components.

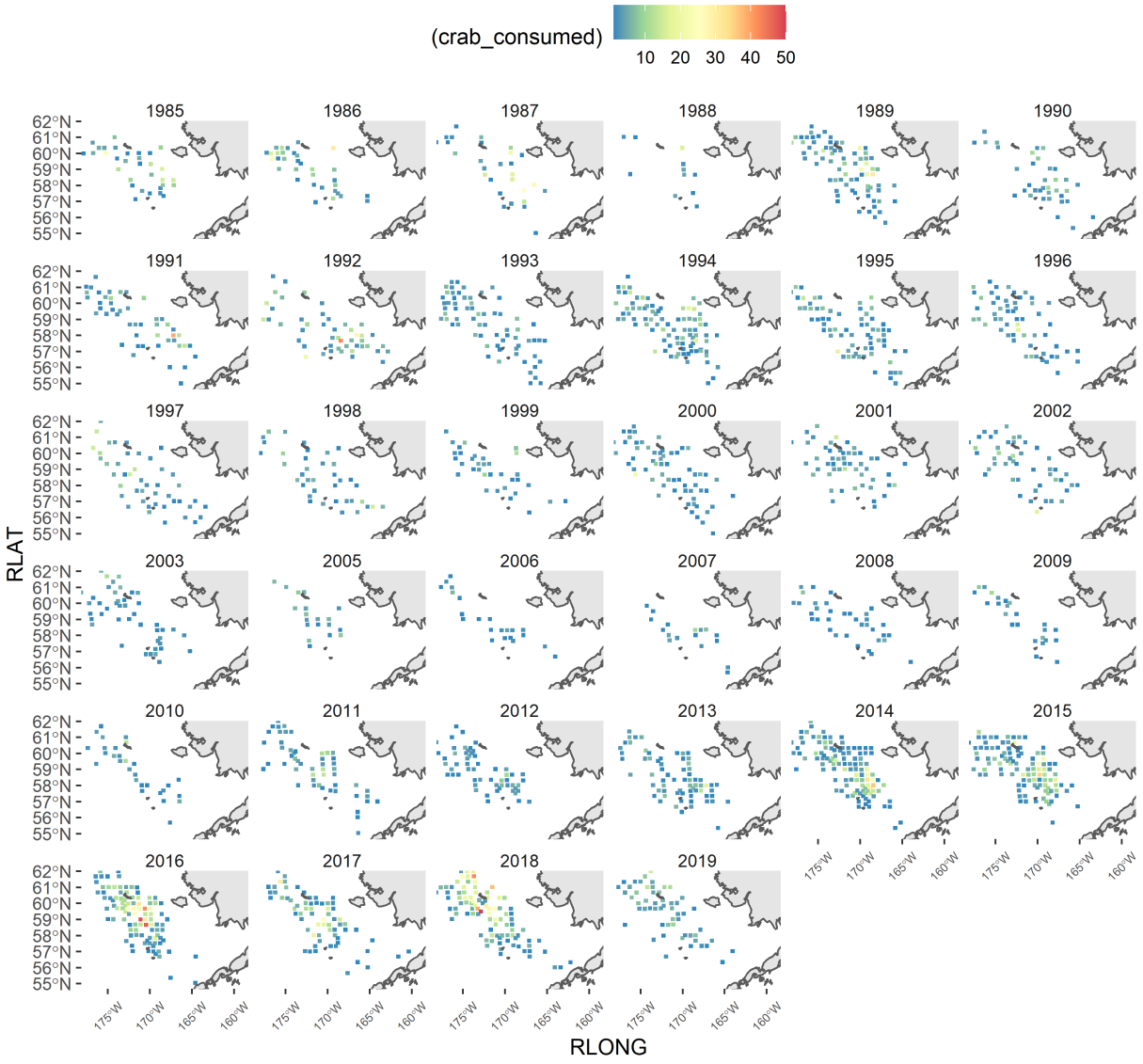

Figure 24: Location and number of crab observed in cod stomachs over time. The are the raw data used to calculate crab consumption by cod and have not been adjusted for sampling effort, but provide background for the spatial distribution of predation over time.

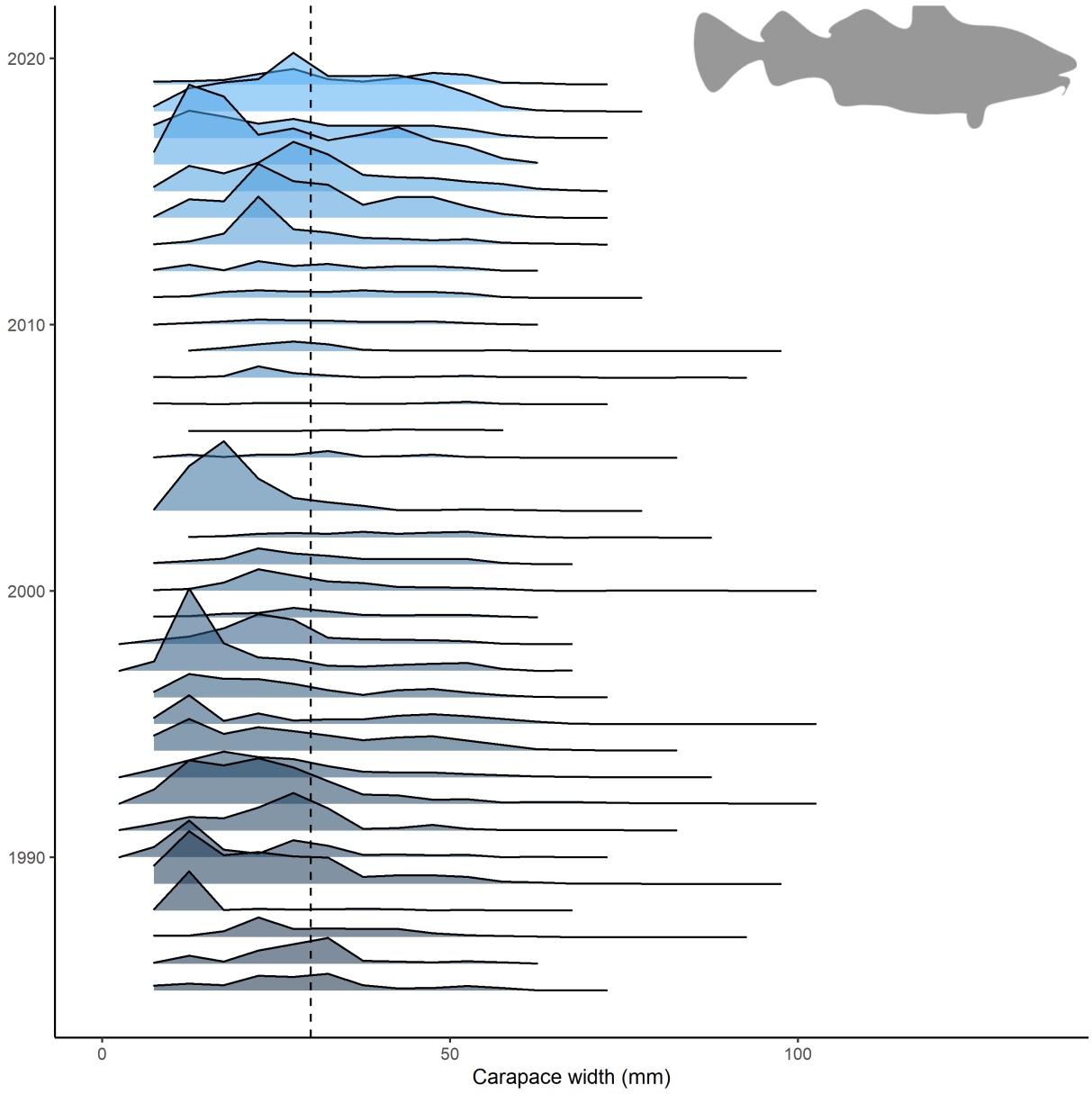

Figure 25: Consumption of crab by Pacifi cod at size over time.

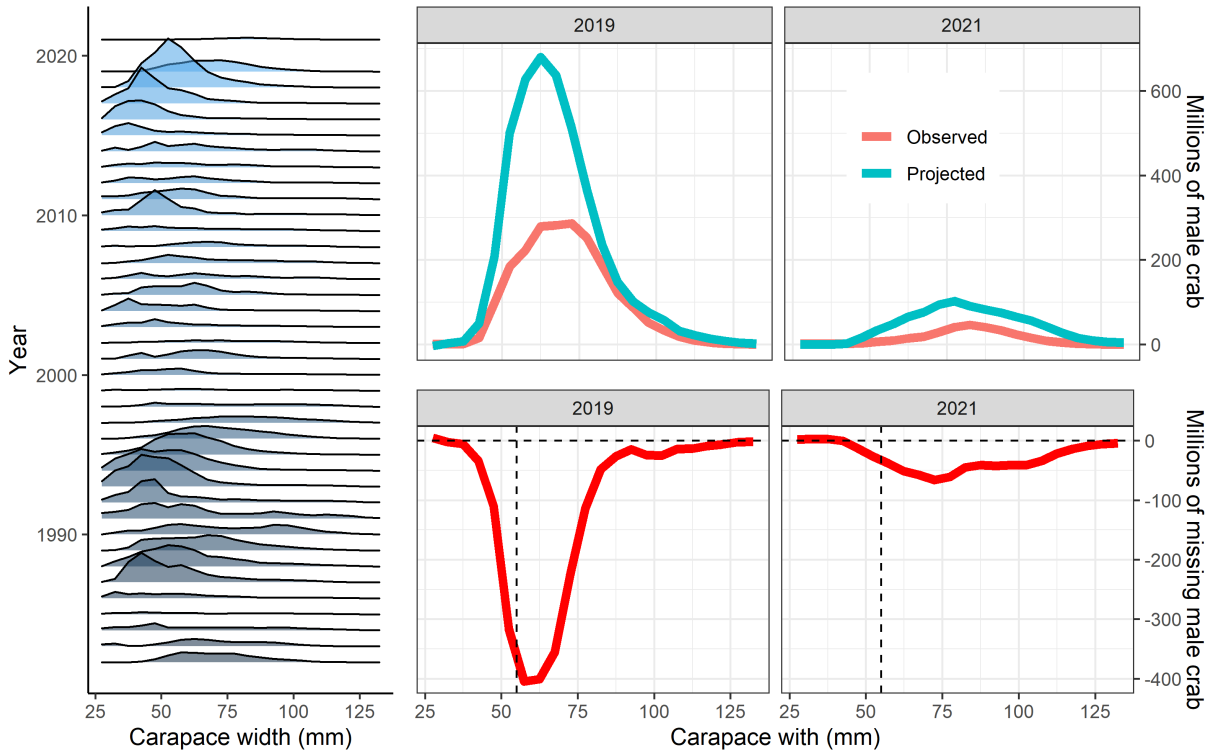

Figure 26: Numbers at size over time of snow crab (left). Observed numbers of crab (red line) in 2019 and 2021 vs. projected numbers of crab from 2018 and 2019 given a mortality equal to 0.27 (the assumed value in the assessment; top left). Numbers of missing crab at size (red line) with the size of crab beneath which cod predate upon (dashed vertical black line).

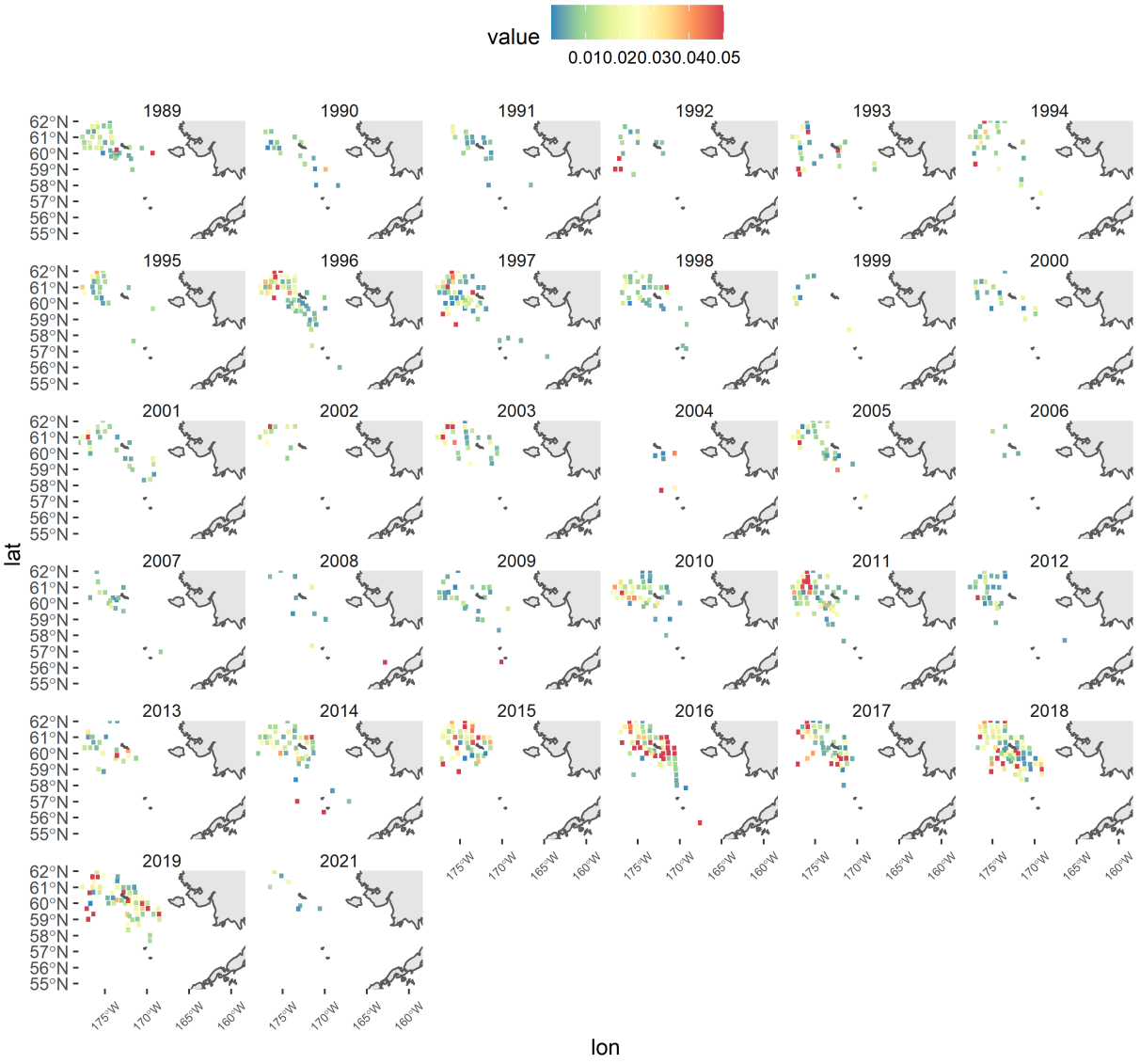

Figure 27: Location and intensity of bitter crab disease over time from visual prevalence observations in the NMFS summer survey.

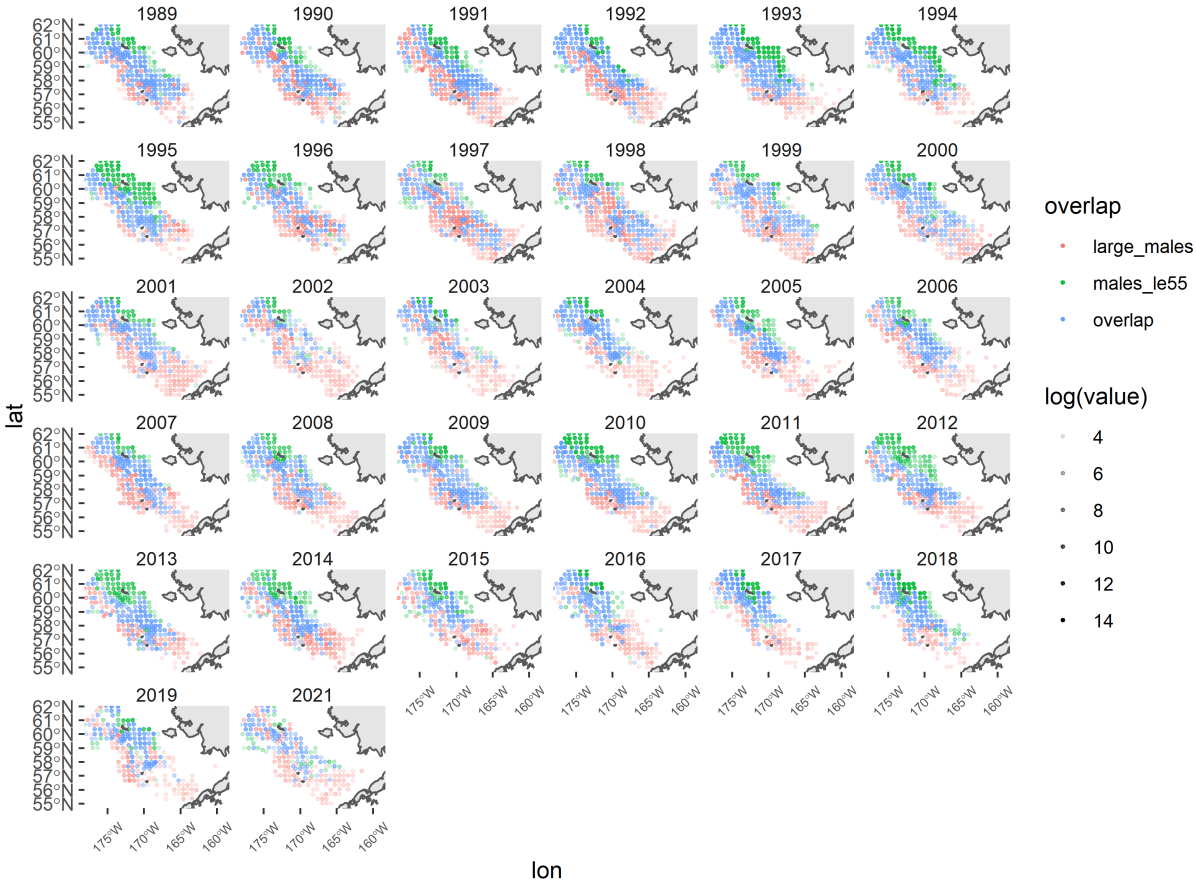

Figure 28: Overlap of large males (>95 mm carapace width) and males smaller than 55 mm carapace width. Opacity of the dot represents the density of crab. Blue represents overlapping distribution. Green and red represent non-overlapping observations of small and large males, respectively.

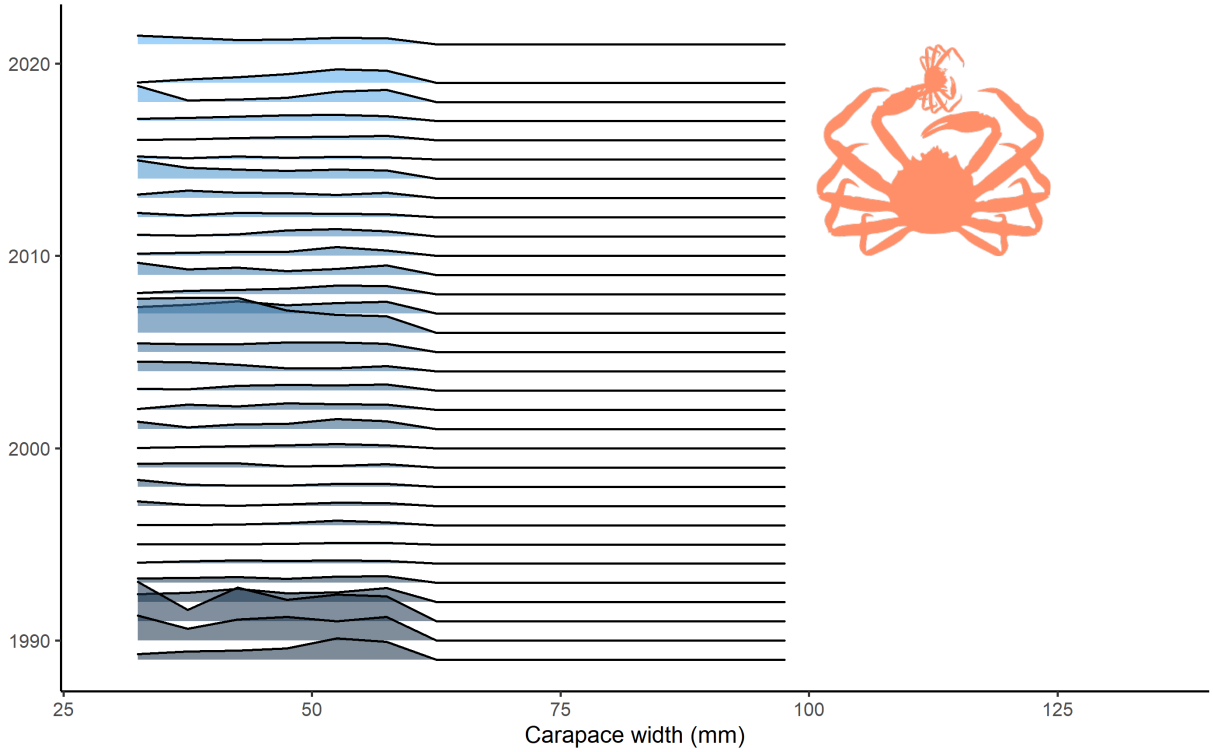

Figure 29: Relative risk at size for cannibalism over time.

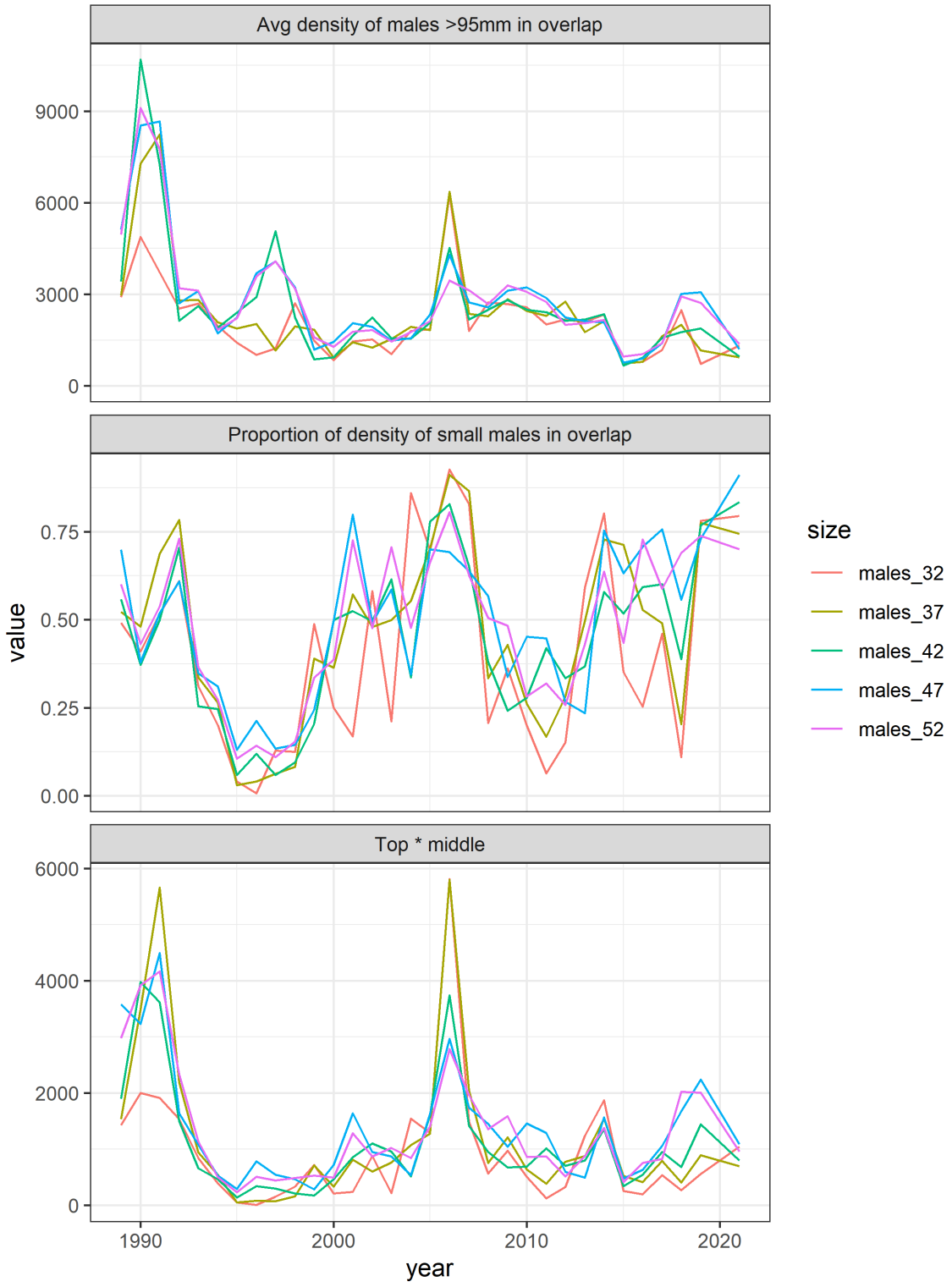

Figure 30: Times series by size of he density of large males in overlapping space (top), the propotion of small males in the overlapping area (middle), and the product of the two (bottom), which is used as an index of cannibalism in the models relating estimated mortality to environmental stressors.

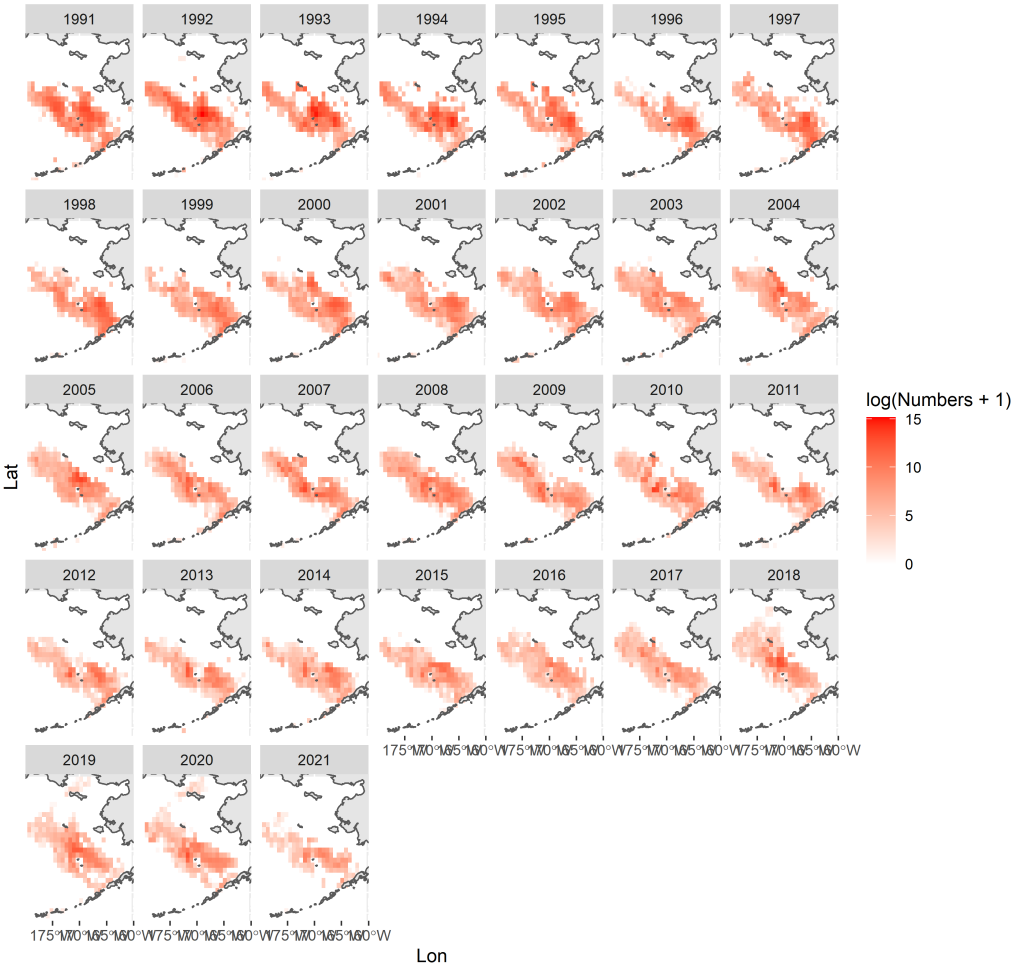

Figure 31: Location and intensity of bycatch of snow crab over time in log space.

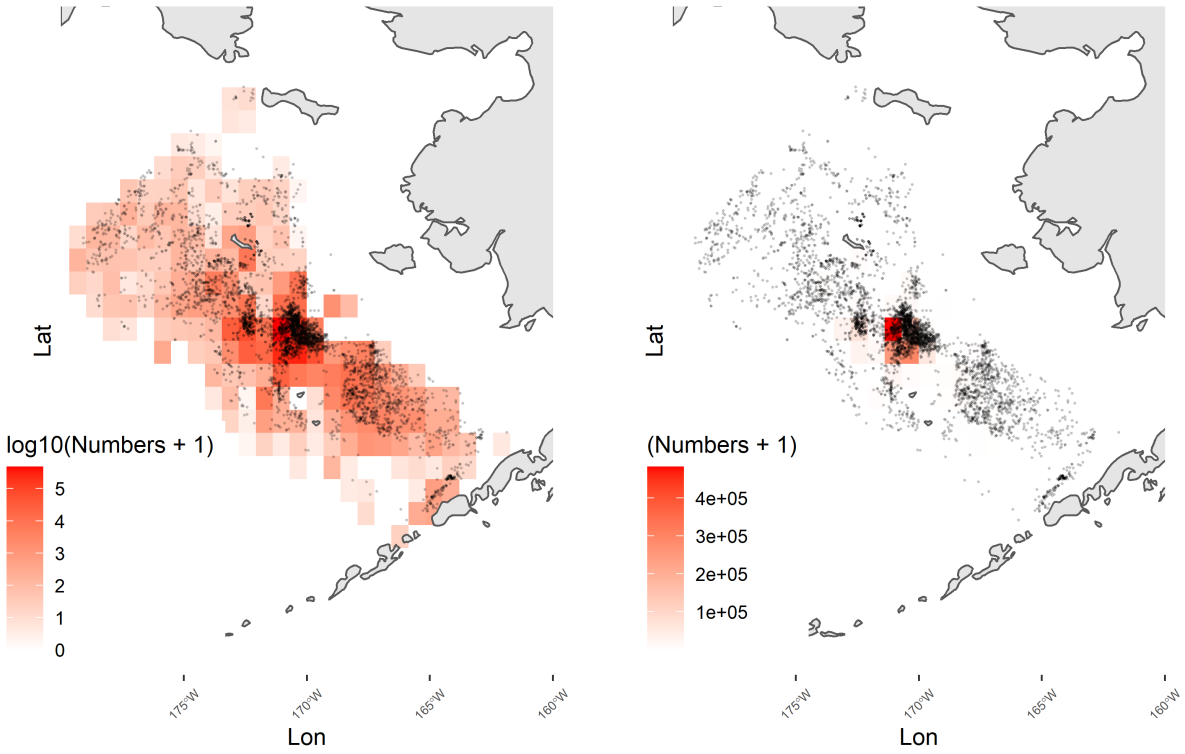

Figure 32: Comparison of location and intensity of bycatch in 2018 for natural and log space.

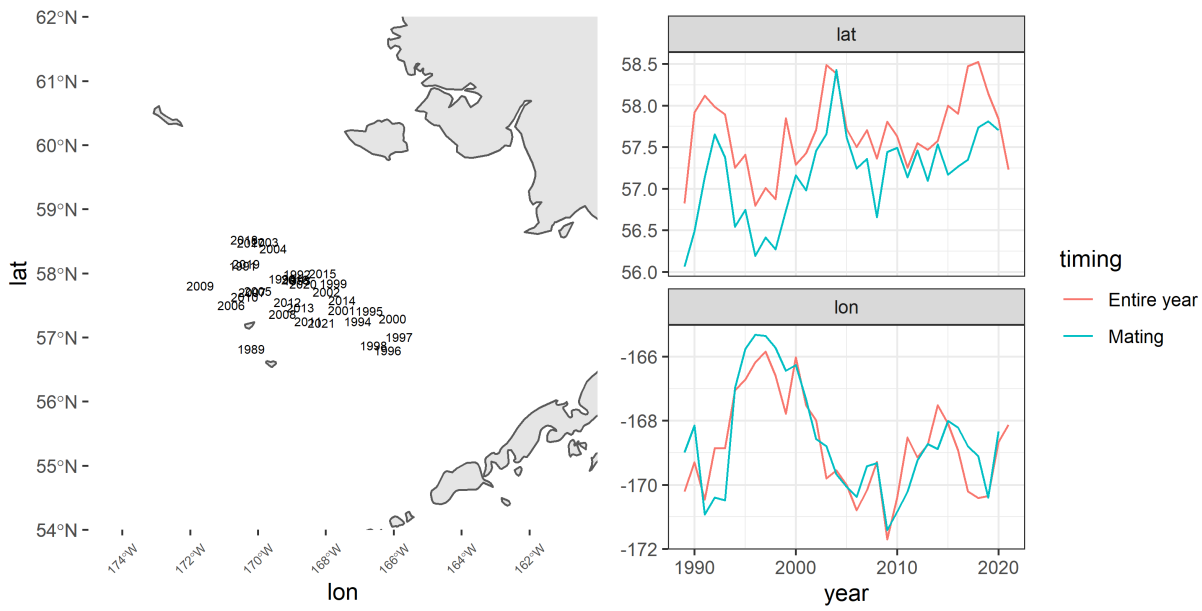

Figure 33: Centroids of bycatch over time calculated over the entire year (left). Centroids broken into time series of latitudinal and longitudinal components calculated over the entire year and during the months December through March which should roughly overlap with mating.

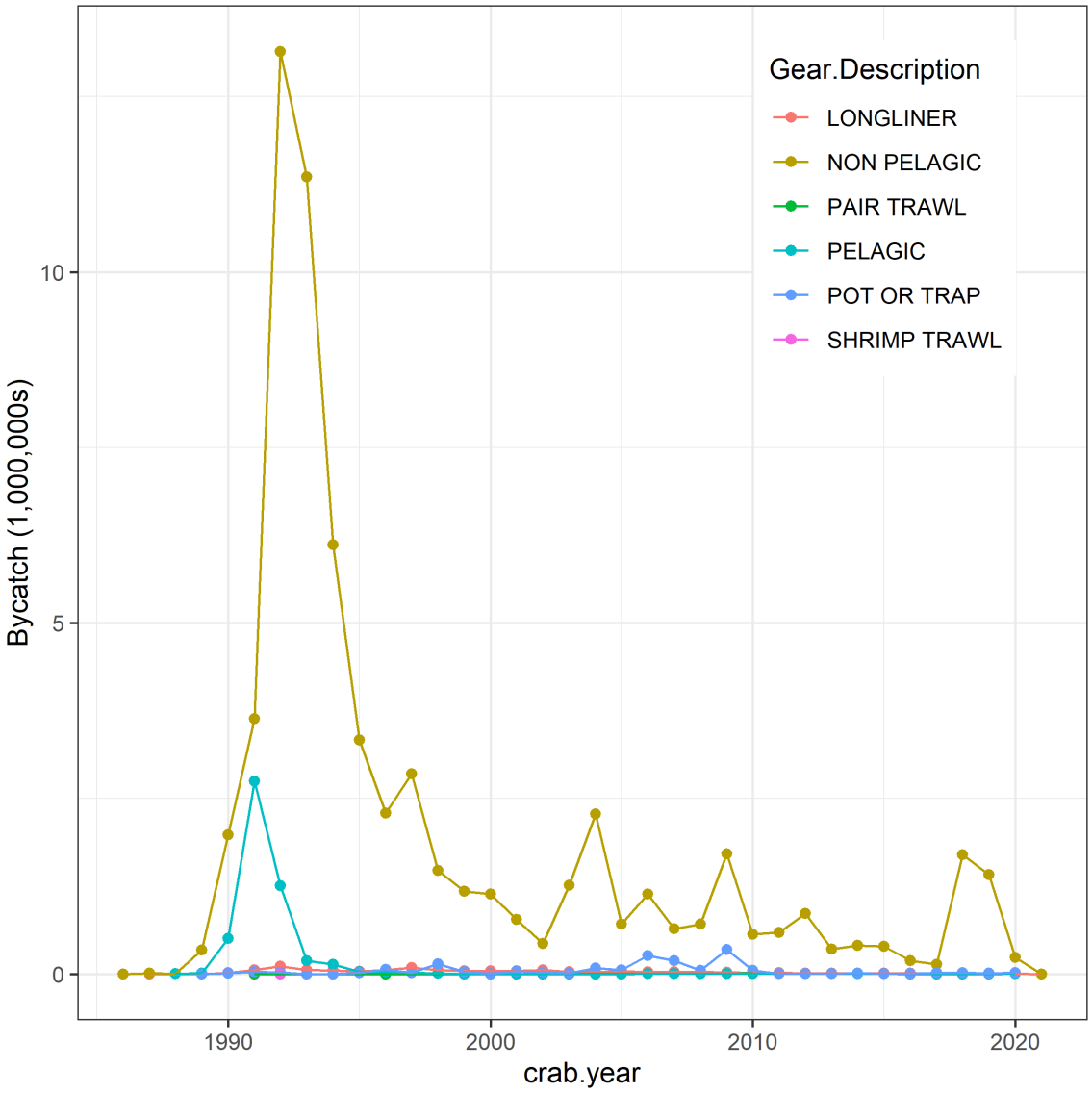

Figure 34: Bycatch by gear types reported from observer programs.

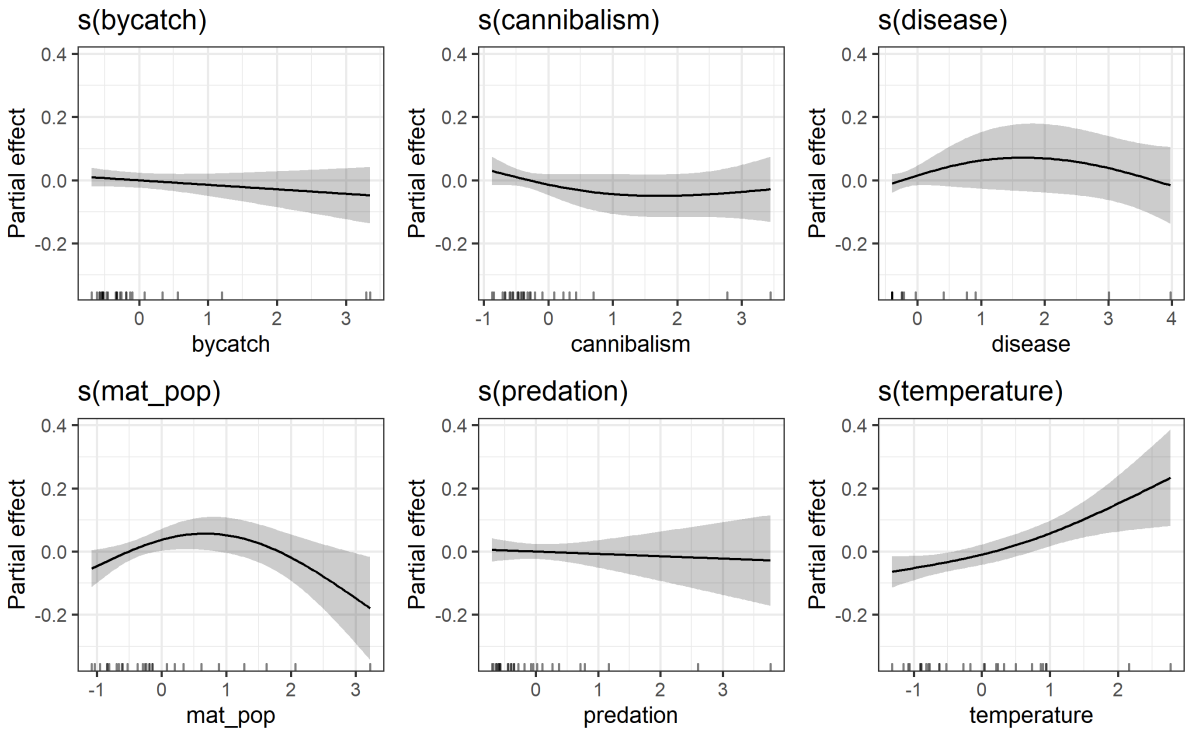

Figure 35: Smooths resulting from the full model estimating the relationship between environmental covariates and immature mortality.

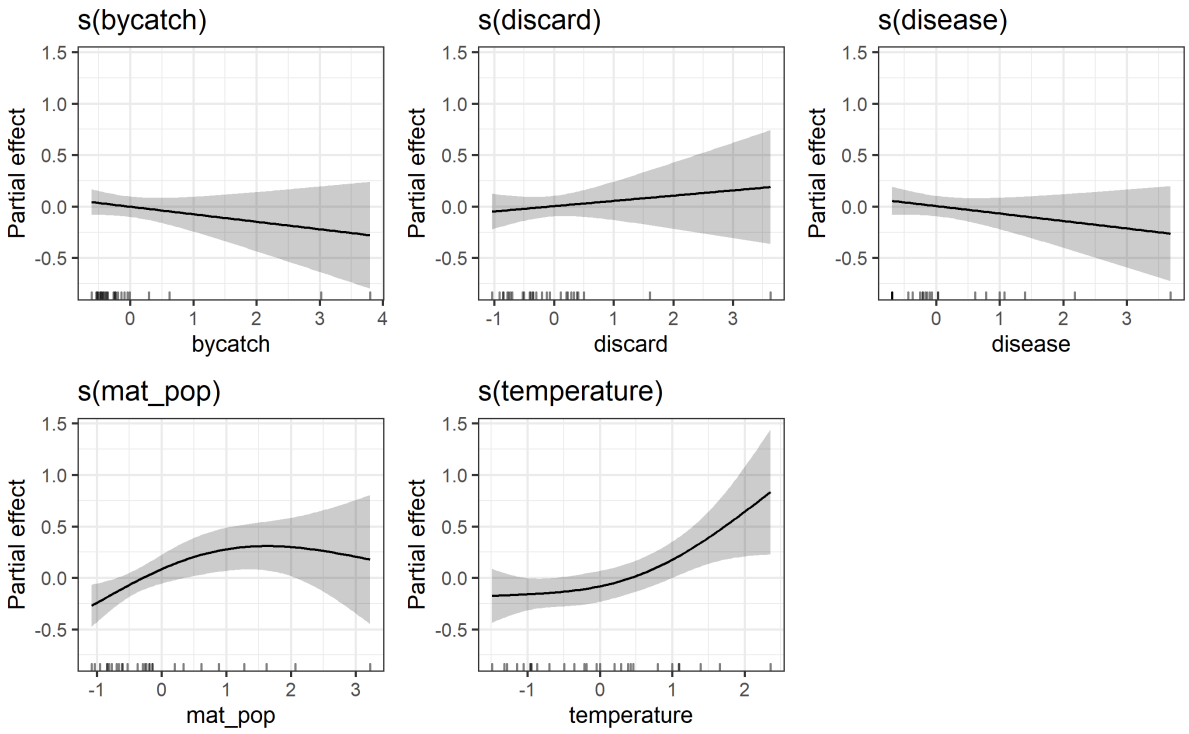

Figure 36: Smooths resulting from the full model estimating the relationship between environmental covariates and mature mortality.

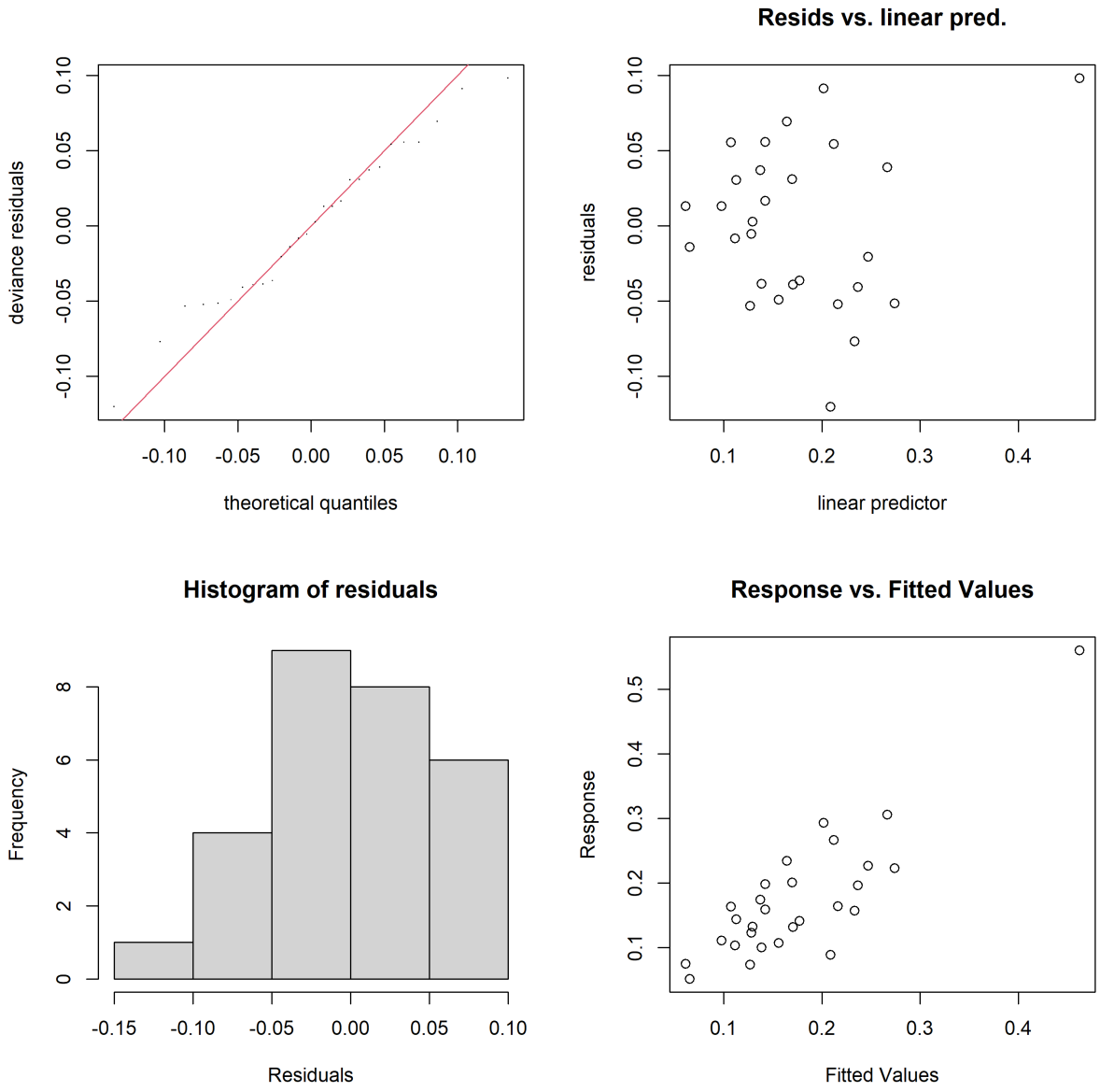

Figure 37: Diagnostic plots for the full models relating immature mortality and environmental stressors.

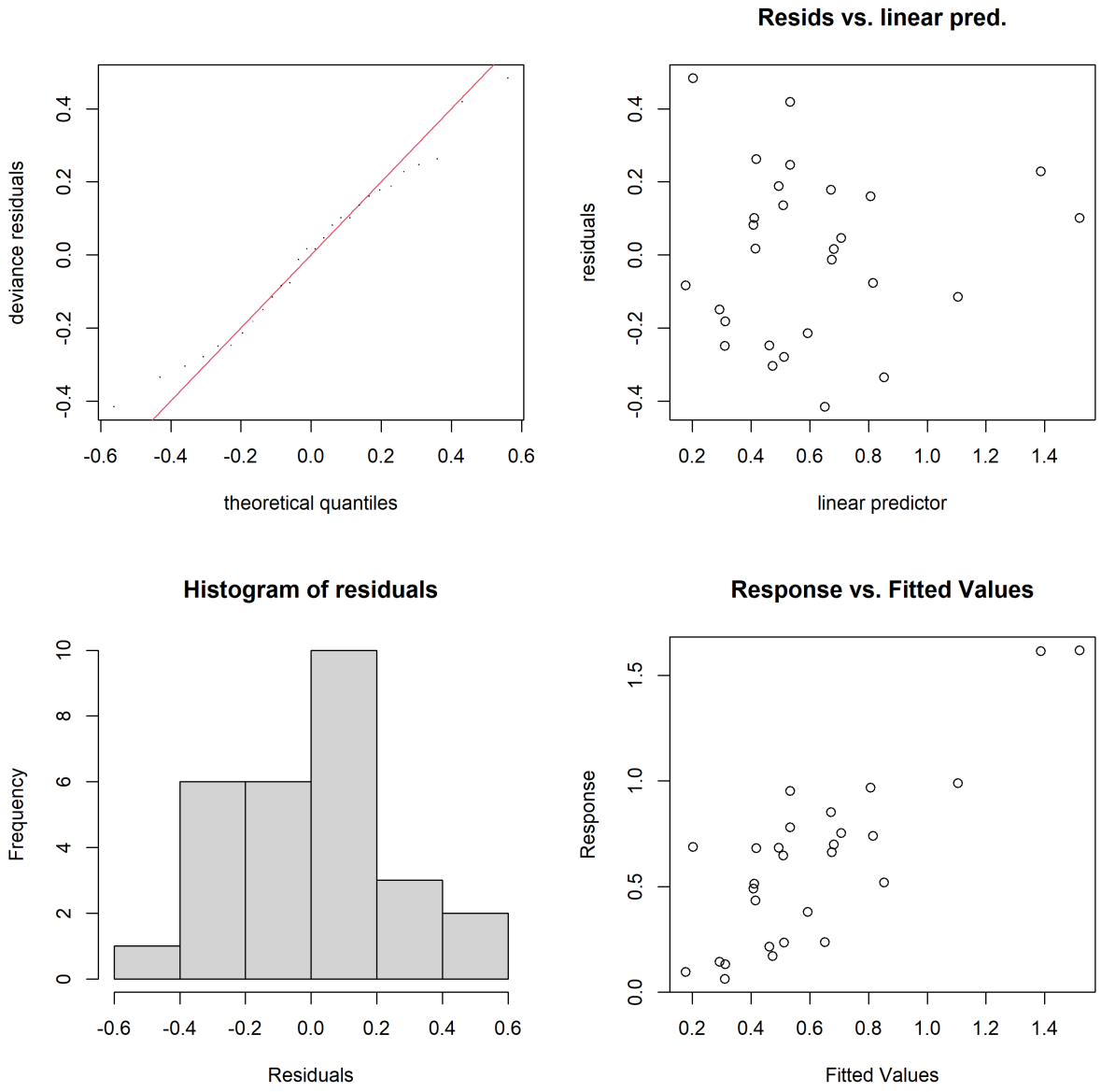

Figure 38: Diagnostic plots for the full models relating mature mortality and environmental stressors

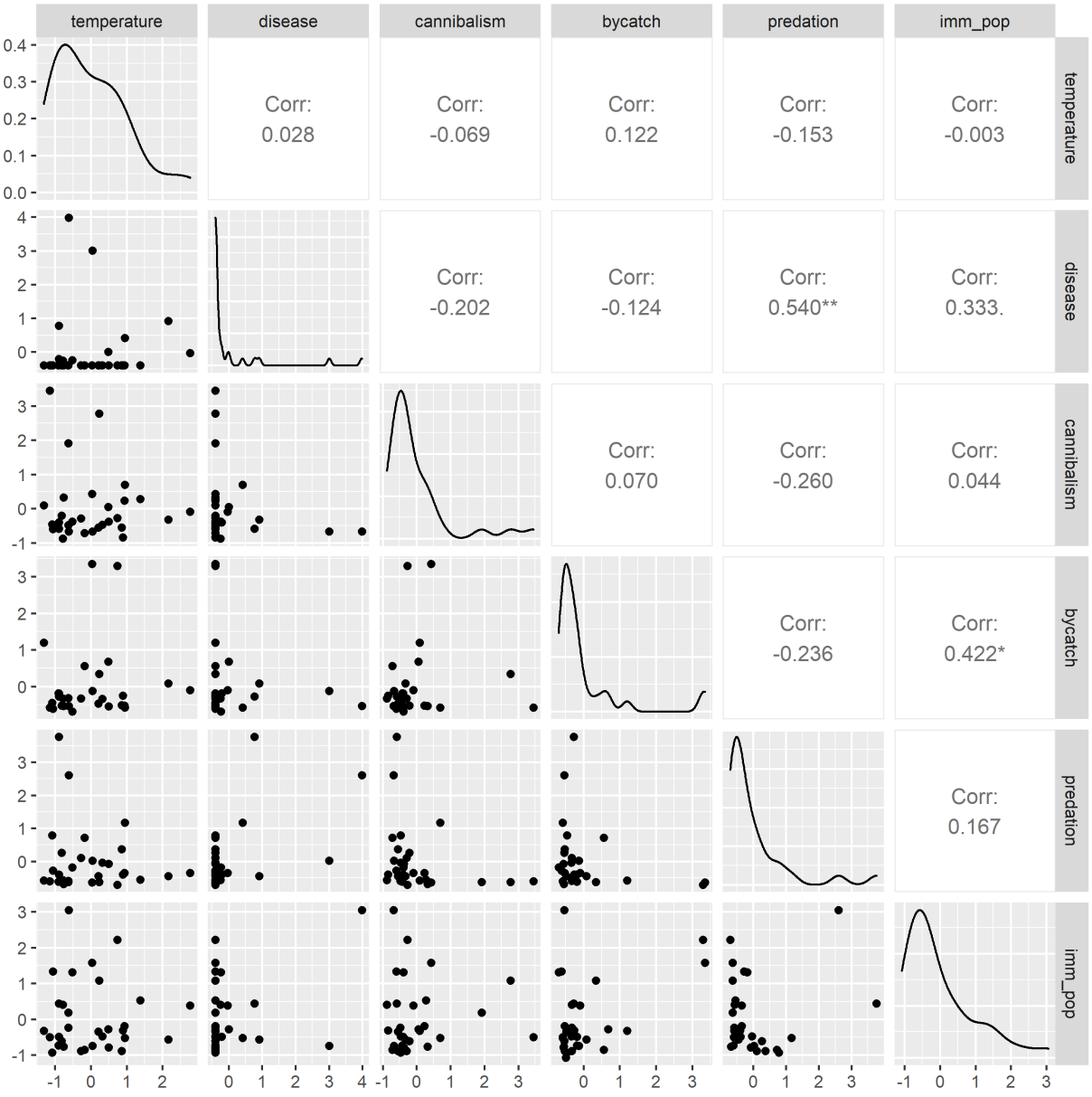

Figure 39: Pairs plots displaying the correlation between covariates for immature crab.

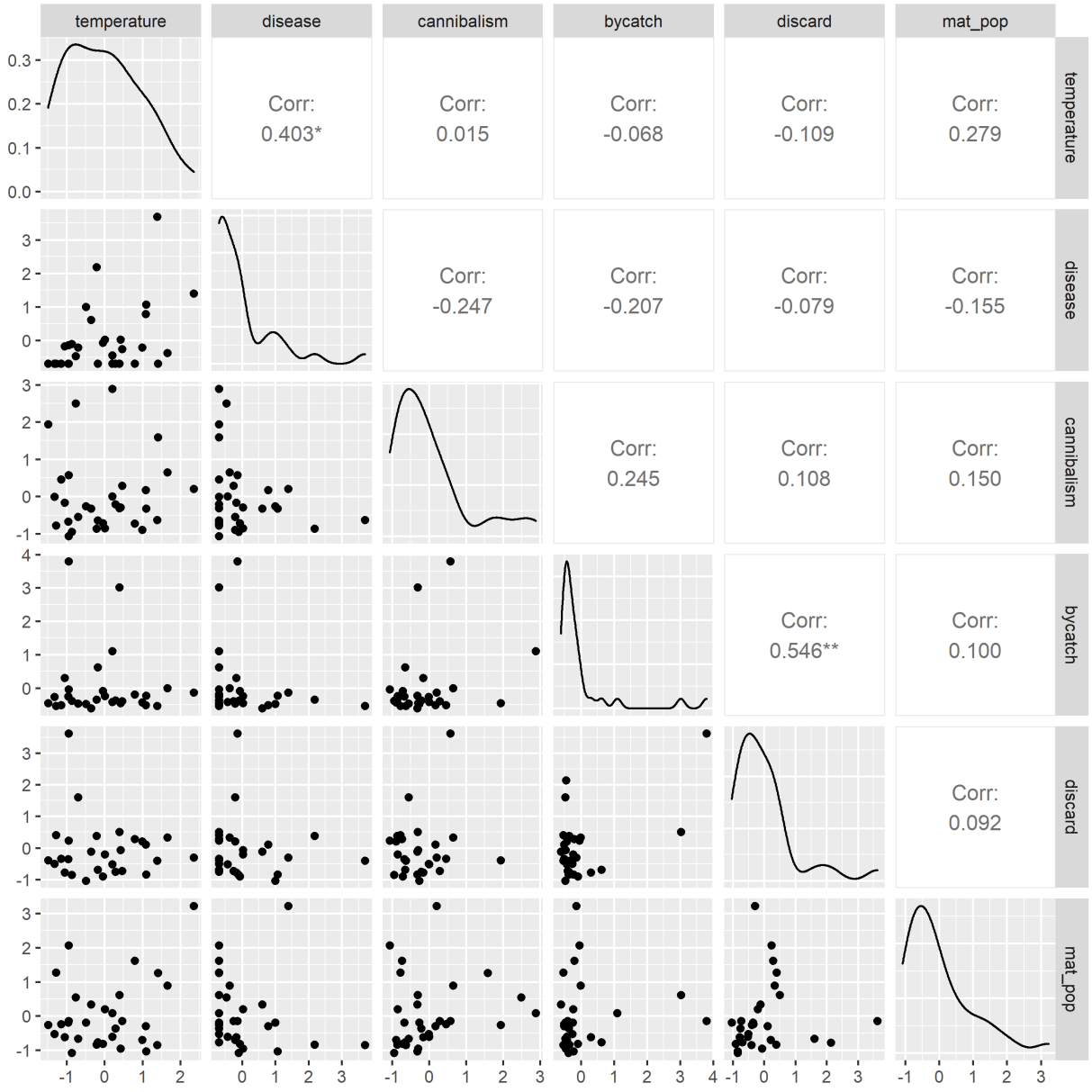

Figure 40: Pairs plots displaying the correlation between covariates for mature crab.

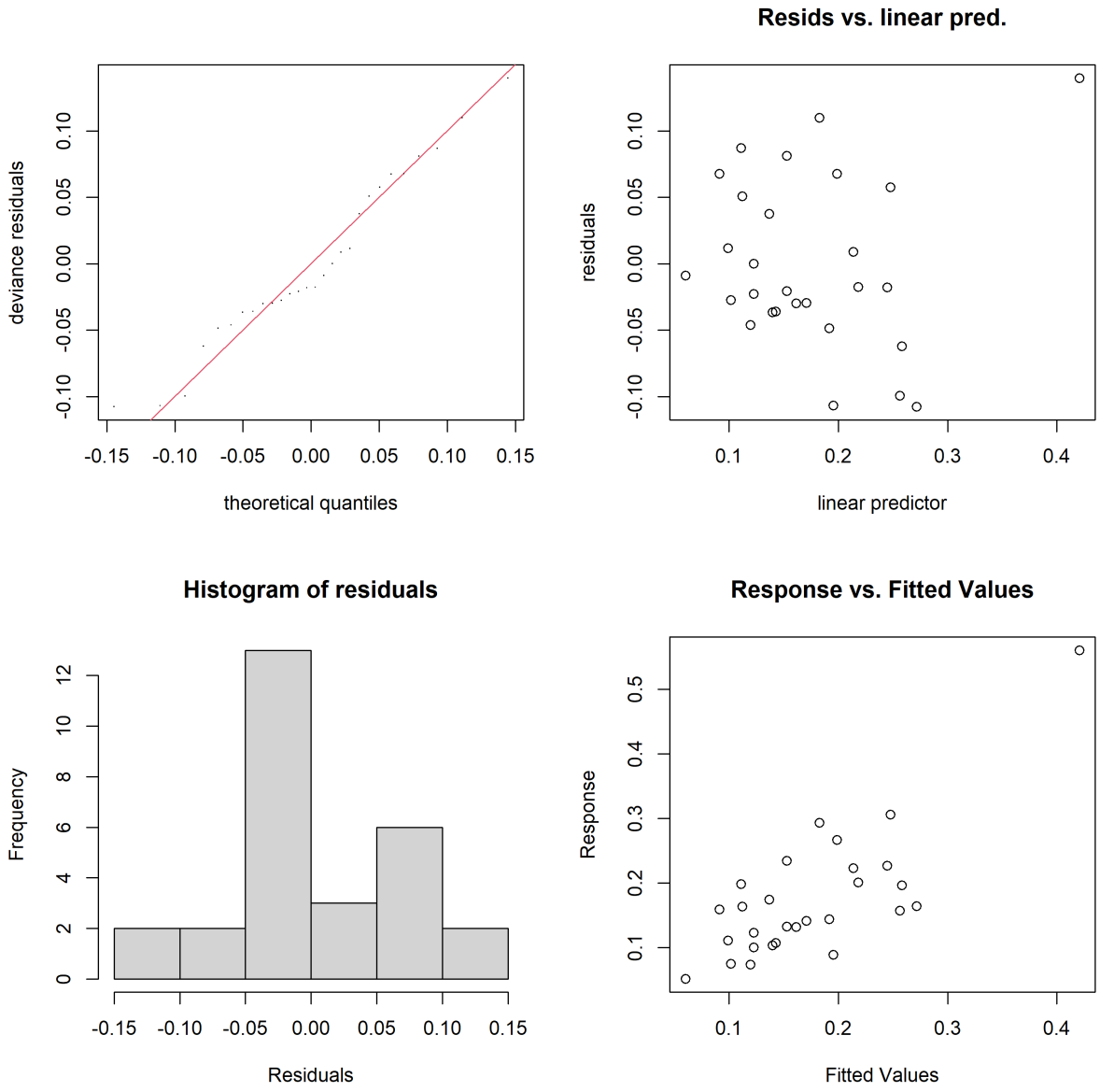

Figure 41: Diagnostic plots for the trimmed models relating immature mortality and environmental stressors.

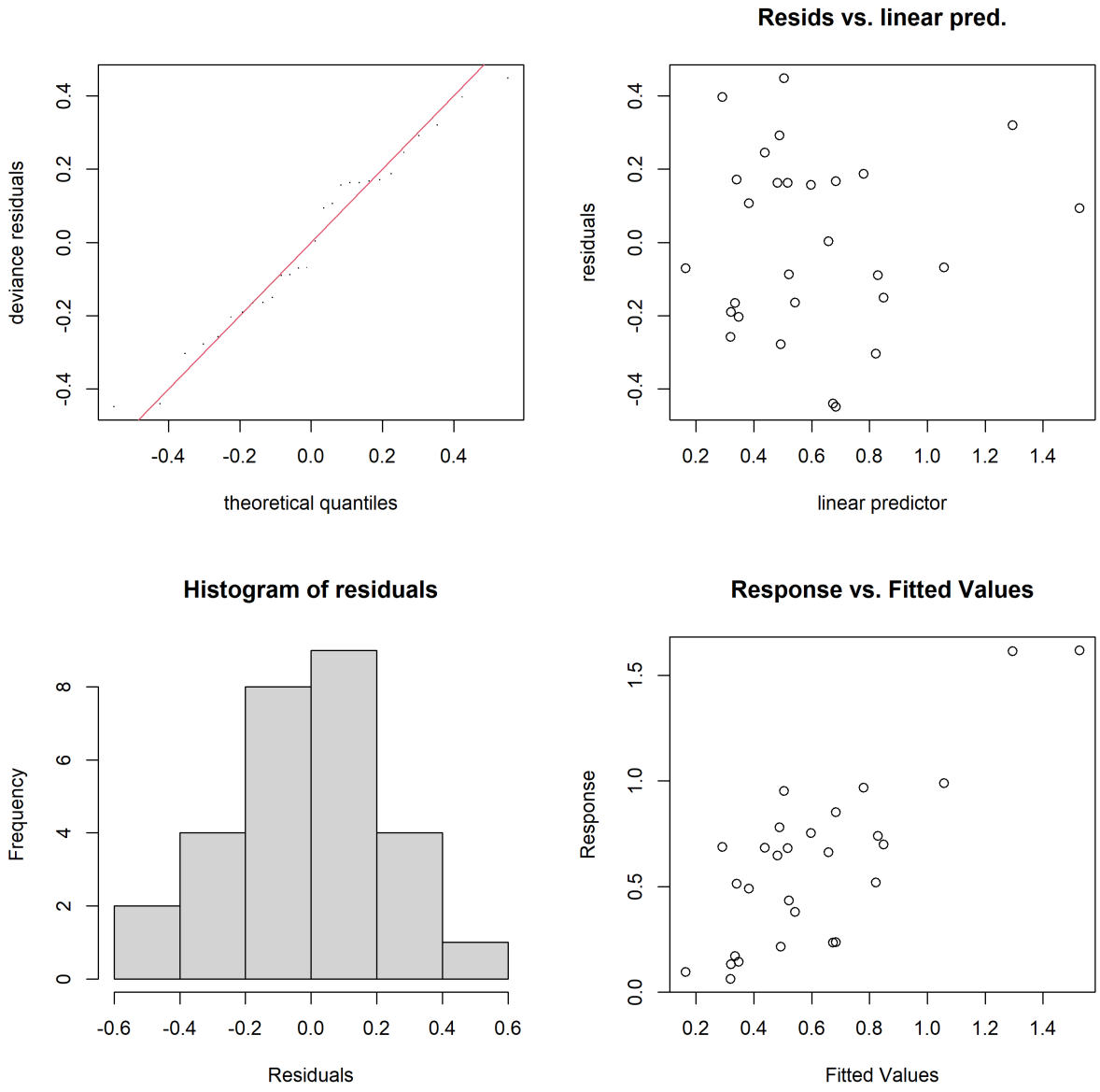

Figure 42: Diagnostic plots for the trimmed models relating mature mortality and environmental stressors.

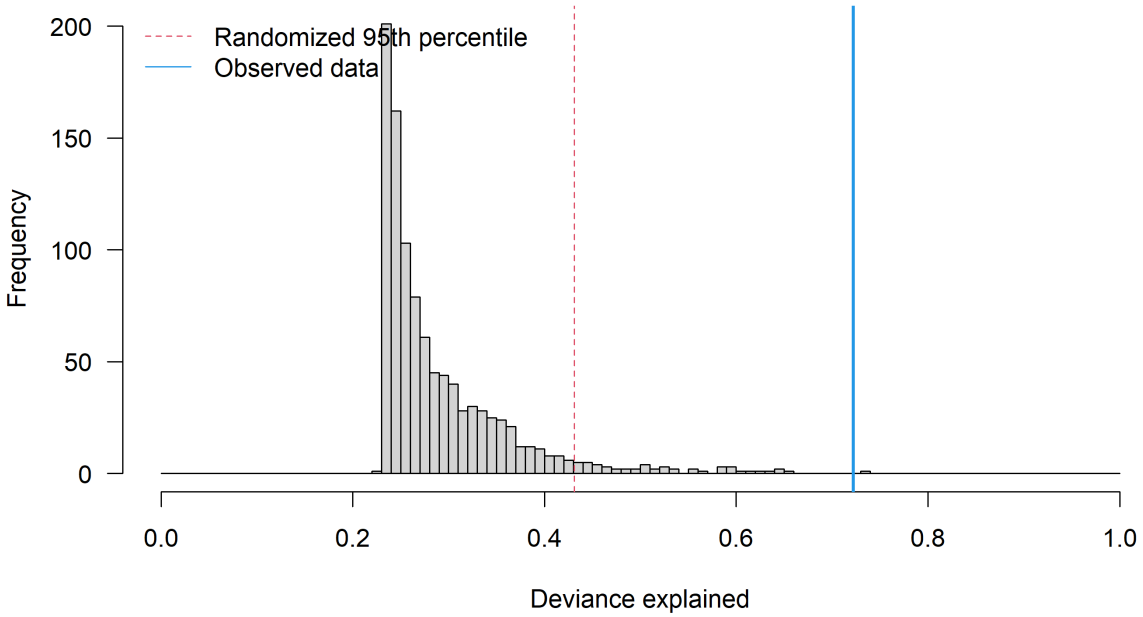

Figure 43: Results of randomization trials for the trimmed models relating estimated immature mortality to environmental stressors. Grey bars represent the number of trials in which the randomized model explained the deviance on the x-axis. Dashed vertical red line represents the 95th quantile of the deviance explained by the randomized trials. Blue line represents the deviance explained with the real data.

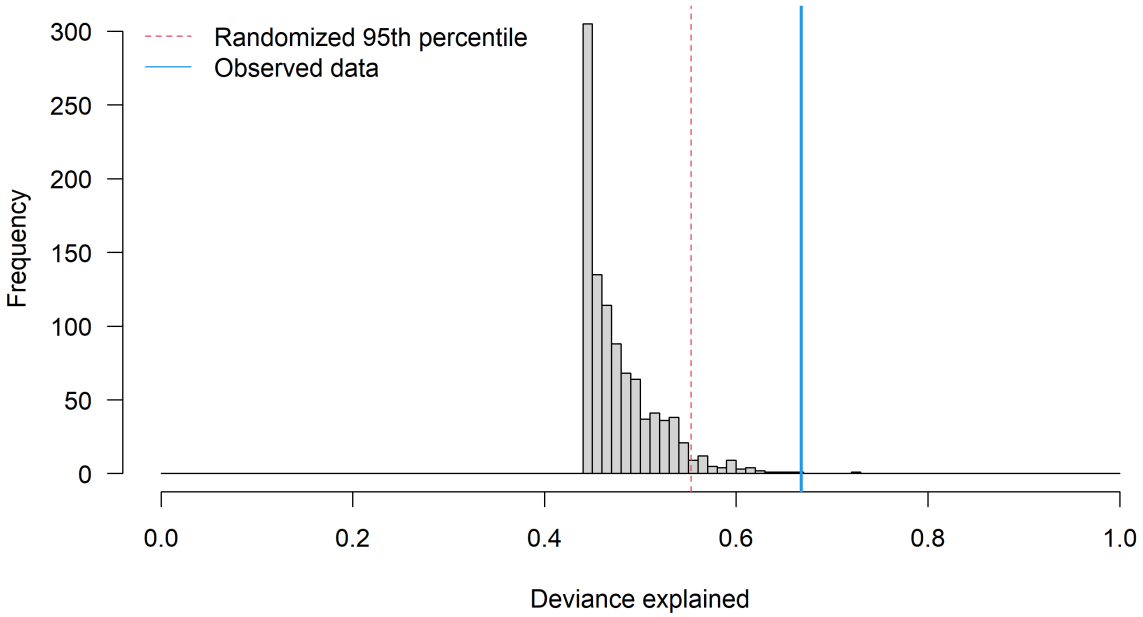

Figure 44: Results of randomization trials for the trimmed models relating estimated immature mortality to environmental stressors. Grey bars represent the number of trials in which the randomized model explained the deviance on the x-axis. Dashed vertical red line represents the 95th quantile of the deviance explained by the randomized trials. Blue line represents the deviance explained with the real data.

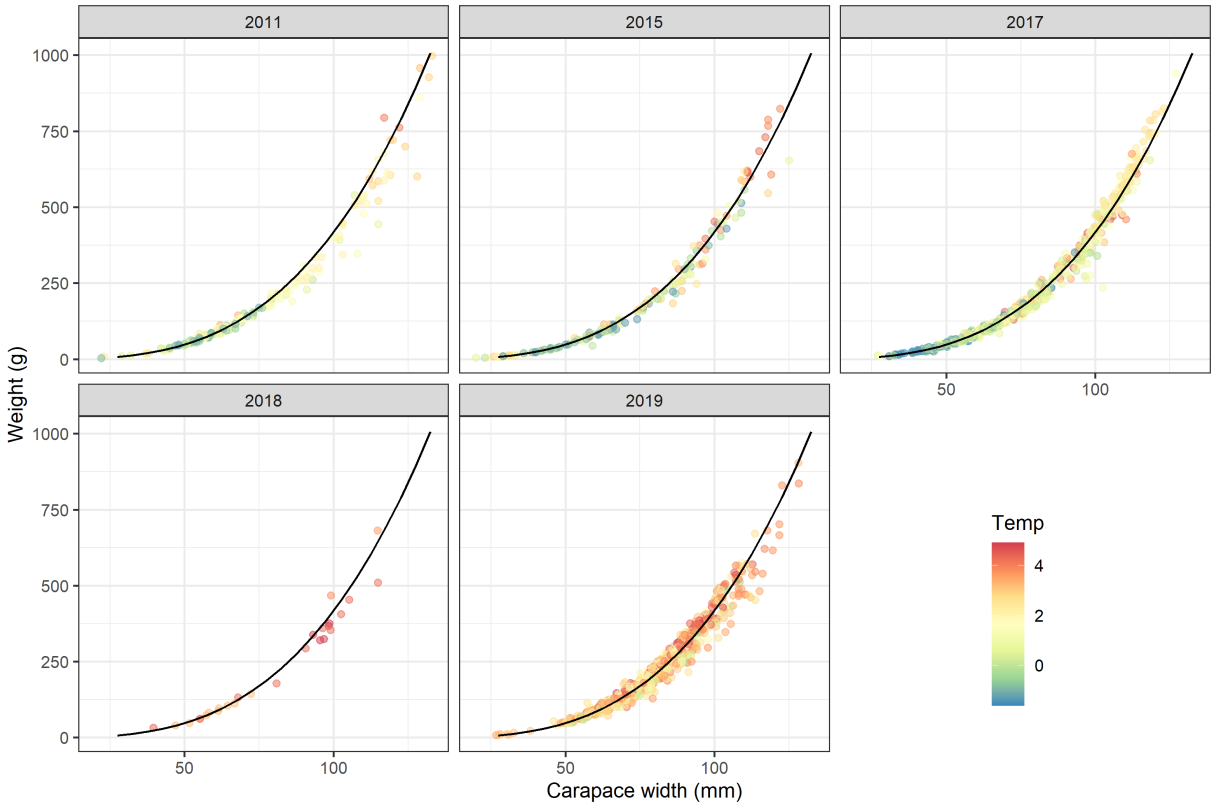

Figure 45: Observed weight at size over time colored by temperature at which the crab was collected.

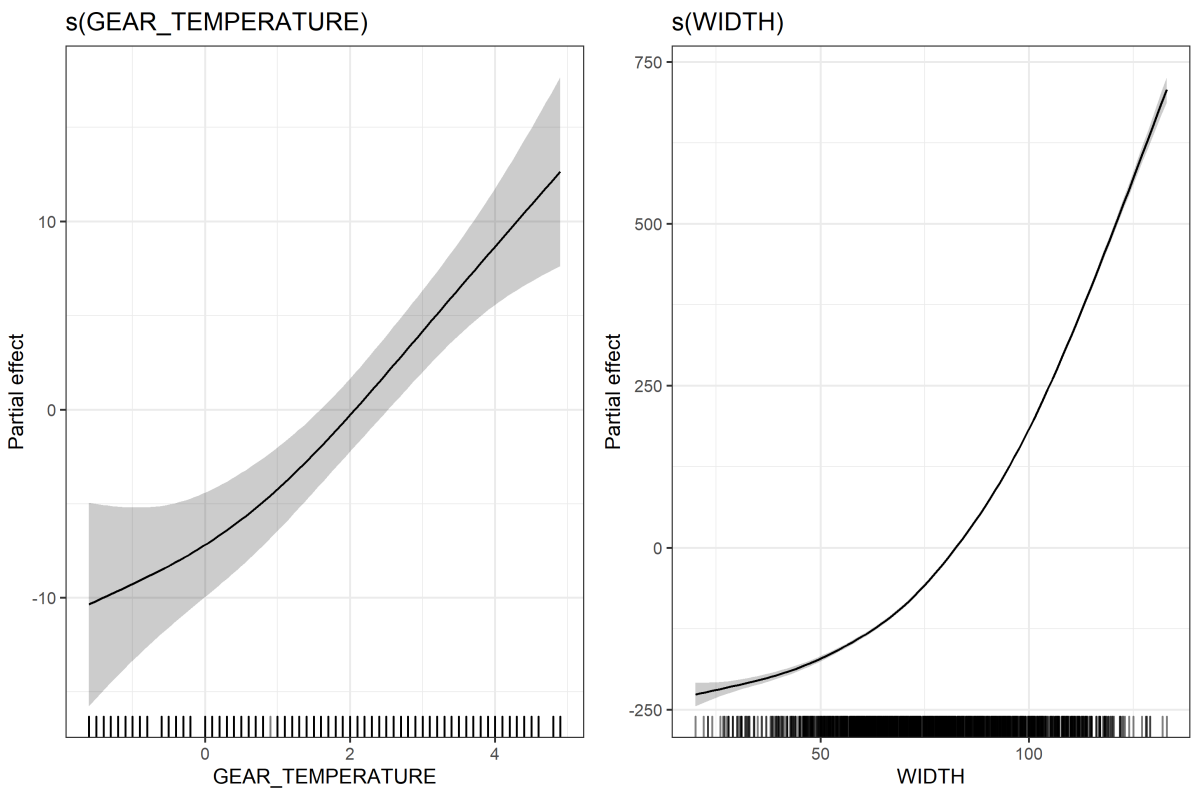

Figure 46: GAM estimated influence of temperature and carapace width on observed weights of crab.