Supplementary materials for 'Canaries of the Arctic: the collapse of eastern Bering Sea snow crab'

Contents

Supplementary materials **1**

Methods overview . . . 1

Population dynamics model . . . 1

Survey selectivity . . . 3

Objective function . . . 3

Penalties and priors . . . 4

Population dynamics model sensitivities . . . 4

Does allowing mortality or catchability to vary over time improve model fits? . . . 4

How well can the model estimate mortality and selectivity with simulated data? . . . 5

How do the assumptions about weighting and priors influence the estimated quantities? . . . 6

Covariate construction . . . 6

Temperature . . . 7

Predation . . . 7

Disease . . . 8

Cannibalism . . . 8

Fisheries data . . . 9

Crab density . . . 9

Generalized additive models . . . 9

Sensitivities to model assumptions in GAMs . . . 11

Error structure . . . 11

Does incorporating the uncertainty in the estimates of mortality change model outcomes? . . . 11

Shape of estimated relationships . . . 12

How could temperature relate to mortality mechanistically? . . . 12

A word on methods . . . 14

Frequently asked questions . . . 14

Are you sure the collapse wasn't a result of cod predation? . . . 14

Are you sure the collapse wasn't a result of trawling? . . . 15

What do crab eat? If they starved, did there appear to be large declines in their prey base? . . . 15

If it was a large mortality event, did you see large numbers of empty carapaces in the survey? 15
Were that many crab really in the eastern Bering Sea to begin with? Was the 'collapse' an artifact of some survey error? 15

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 on 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 prevalence, cannibalism, and crab density from the NMFS survey to test as covariates in GAMs. Cod predation indices were developed using stomach content data collected on NMFS surveys in addition to cod size composition and abundances. Indices for fishery related effects were collated from fisheries statistics from the Alaska Department of Fish and Game and also included in the GAMs.

Ecological detective work in the marine environment is hampered by the difficulty of observation and this is particularly so on the eastern Bering Sea shelf. The waters in which snow crab reside range from 50-200 meters deep and are seasonally covered by ice, making data collection only feasible in the summer. Consequently, the survey based portions of our analyses are derived from a yearly snapshot of the population over an approximately 30 year period. Each of the hypotheses explored here clearly result in some mortality. We know that millions of crab are eaten by cod every year, the directed and bycatch fisheries kill crab, larger crab eat smaller crab, and crab die from bitter crab disease each year. The goal of our analysis is to place each of these processes in a historical context to try to understand the relative impact of each and what was different about the recent collapse. More than one way exists to analyse the available data on this issue. Below we describe our approach, including a description of each of the components of our analysis, a discussion of the rationale behind our modeling decisions, and sensitivities and simulation tests of our models, all of which provide what we think is sound reasoning for our analysis.

Population dynamics model

The population dynamics model presented here incorporated 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. 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 (Tamone et al., 2005). 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 5). 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 between pre- and post-molt carapace width (Figure 6), ($\hat{W}^{pre}_{s,w}$ 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 6).

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

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

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

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

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 6) 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, molting, 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 *Nephrops* trawls (a small trawl net designed to maintain bottom contact), operated 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 7). 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 8). The selectivity of large crab determines the estimated scale of the population in a population dynamics model, but the information we have on selectivity of large crab 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 (representing the fit of the model to the data) and penalty components (which incorporate constraints in the fitting based on prior information) 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 50.

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 in the simulation analyses aimed at understanding the estimability of mortality and catchability) using normal likelihoods on the second differences of the vectors. Normal priors were also placed on the mean value of natural mortality and catchability and the deviation of the estimated mortality from that mean. 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. The normal priors were of the form:

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

P_x was the contribution to the objective function of the penalty associated with model estimate x, λ_x was any additional weighting applied to the component, $\hat{I}_{x,y}$ was the predicted value of population process I relevant to penalty x during year y, $\mathrm{I}_{x,y}$ was the prior value of process I relevant to penalty x during year y and $\mathrm{CV}_{x,y}$ was the input coefficient of variation for penalty x during year y.

An example of the way in which these equations were implemented can be seen in lines 132-218 of 'snow_down.TPL' in our github repo 'snow_down/models/model_vary_m'.

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 penalties in 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 on model output of including time-variation in mortality and catchability 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 9 & Figure 10). 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 9). 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 11 to Figure 12). 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 by 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 13 to Figure 14). 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 15) 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 16), but larger changes in estimated mortalities and catchabilities (Figure 17). 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 and relatively tight priors on the mean mortality given outside information to support an average mortality value based on longevity. 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).

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 (Zacher et al., 2022). The fisheries-dependent observer data provide spatio-temporal information on bycatch (AKFIN, 2022). The main text notes that more than 10 billion crab have gone missing since 2018. This number is derived from the input total numbers observed in the survey to the assessment, which decreased from 11.7 billion animals in 2018 to 940 million animals in 2021. However, this figure does not account for the selectivity of the survey gear and includes both sexes. If survey selectivity is accounted for, the number of missing crab increases dramatically, with the most recent assessment estimating a decline from ~47 billion in 2017 to 2.58 billion in 2022. Regardless of the metric used, the number of crab missing from the Bering Sea survey was exceptionally large.

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 estimated 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 experienced by immature crab can be different than those 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 in a given year. The size at which more than half of the population is mature changes by year, depending on recruitment dynamics and other demographic processes (Figure 18). After the survey data were split at the 50% at maturity size, time series of maturity-specific environmental stressors (Figure 19) 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, 2008). The cold pool, a mass of water <2 degrees Celsius, can act 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 20). As the cold pool changes from year to year, so does the spatial distribution of snow crab (Figure 21). The ontogenetic migration of snow crab results in crab of different sizes and maturity states experiencing different temperatures in a given year (Figure 22). 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 23).

Predation

Pacific cod (*Gadus macrocephalus*) are the most important predator of snow crab based on stomach content data collected in the NMFS bottom trawl survey (Long and Livingston, 1998), with 16.5% of cod stomachs containing snow crab (Burgos et al., 2010). Crab ranging from 8-57 mm carapace width constitute 95% of the crab consumed by cod in the Bering Sea, but crab up to 106 mm carapace have been observed in cod stomachs (Burgos et al., 2010). An index of summer daily consumption (tons/day) of snow crab between 30-95mm carapace width eaten by Pacific cod in the eastern Bering Sea was developed using cod stomach content data from the survey to estimate the proportion by weight of crab in cod diets and the size composition of crab by carapace width of prey found in cod stomachs, stratified by year, survey stratum, and cod length (collection and analysis methods described in Livingston et al. 2017). Cod total consumption rate (metabolic demand) was calculated using a cod bioenergetics model (Holsman and Aydin 2015) to estimate laboratory-measured maximum consumption rates adjusted for bottom water temperatures and cod abundance-at-length measured at each haul location (following methods described in Barbeaux et al. 2020), and summed to an eastern Bering Sea ecosystem-wide total.

Changes in the cold pool can alter 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 24 & Figure 25). This increased interaction coincided with increased numbers of crab consumed by cod in the last several years (Figure 4). The estimated number of cod greater than 50 cm was also near all-time highs around the period during which crab collapsed (Figure 26). 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 or increases in abundance of large cod. 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. Predation indices were calculated for crab by year $P_{m,y}$ by calculating the ratio of the extrapolated biomass of crab consumed by cod to the estimated biomass of crab, $N_{y,m,s} * w_s$:

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

The exact amount of crab eaten cannot be calculated from the available diet data because they are a snapshot of 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. The index of consumption described above incorporates the most available data on cod predation, but some strong assumptions are made (e.g. summer diet is representative of the entire year). As a sensitivity to these assumptions, we also tested the ratio of the number of cod greater than 50 cm to crab abundance in a given year as an alternative index of predation in the GAMs. Ultimately, changing the index of predation did not impact the results of the fitting of the GAMs; temperature and mature population size were still the only significant covariates and the estimated shapes of relationships and deviance explained were very similar between models with the different predation indices. Consequently, the models presented in the main text use the index of consumption as the predation index because it uses the most available information on cod predation (i.e. stomach contents and the abundance and size composition of cod).

Disease

Bitter crab syndrome is a fatal disease in snow crab caused by a parasitic dinoflagellate (Meyers et al. 1996). 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. Diseased crab are visually detected by a pink-orange discoloration of the carapace and opaque hemolymph. 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 for the respective maturity states (Figure 19).

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 smaller than 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). Indices of cannibalism were only included in the immature models given laboratory observations indicate cannibalism is rare among crab of similar sizes, though molting crab can be vulnerable.

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 male crab are often discarded and all females are 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 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. Some fraction of the mortality imposed by non-directed fleets is likely unobserved due to crab being struck by the gear and not captured. Consequently, 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.

Crab density

The numbers of crab estimated from the population dynamics models were also used as covariates in the GAMs. Changing densities of crab could capture aspects of intraspecific competition not captured in other covariates. Each respective model of mortality incorporates the population size of the corresponding maturity state given their spatial co-occurence. Immature mortality also incorporates mature population size because crab are thought to move more extensively after maturing in the pursuit of mates, which suggests that their overlap with the immature portion of the population could be larger than the snapshot the survey provides. This increased overlap could result in impacts on mortality, hence the inclusion of mature population size in the immature mortality models.

Generalized additive models

Generalized additive models (GAMs) were used in the R programming language (package mgcv; Wood, 2011) to relate changes in estimated mortality by maturity state and year, $m_{m,y}$ to environmental covariates by maturity state and year, $\phi_{m,y}$, because of their flexibility in fitting potential non-linear relationships. Models were first fitted in which all potential 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)	0.6035	0.0439	13.7531	< 0.0001
B. smooth terms	edf	Ref.df	F-value	p-value
s(temperature)	2.0477	2.4391	3.6011	0.0334
s(disease)	1.0000	1.0000	0.5203	0.4786
s(discard)	1.0000	1.0000	1.6726	0.2100
s(bycatch)	1.0000	1.0000	0.9665	0.3367
s(mat_pop)	1.8919	1.9799	6.7981	0.0086
s(predation)	1.0000	1.0000	2.9442	0.1009

Table 1: GAM output for full model predicting mature mortality. Deviance explained = 72.04 %

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	0.1719	0.0105	16.4407	< 0.0001
B. smooth terms	edf	Ref.df	F-value	p-value
s(disease)	1.0000	1.0000	1.5639	0.2263
s(temperature)	1.8837	1.9841	10.7801	0.0006
s(mat_pop)	2.0000	2.0000	13.2472	0.0002
s(imm_pop)	1.6241	1.8564	1.6100	0.1681
s(predation)	1.0000	1.0000	0.1759	0.6796
s(bycatch)	1.0000	1.0000	0.0006	0.9808
s(cannibalism)	1.7298	1.9252	2.5843	0.1385

Table 2: GAM output for full model predicting immature mortality. Deviance explained = 77.6 %

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 and Figure 45 for a larger version of figure 2).

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	0.6035	0.0467	12.9274	< 0.0001
B. smooth terms	edf	Ref.df	F-value	p-value
s(temperature)	1.9300	2.3189	4.6542	0.0183
s(mat_pop)	1.8887	1.9763	7.8385	0.0019

Table 3: GAM output for trimmed model predicting mature mortality. Deviance explained = 62.16 %

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	0.1719	0.0121	14.1816	< 0.0001
B. smooth terms	edf	Ref.df	F-value	p-value
s(temperature)	1.7797	1.9506	9.5779	0.0006
s(mat_pop)	1.9813	1.9988	7.4968	0.0031

Table 4: GAM output for trimmed model predicting immature mortality. Deviance explained = 59.5 %

Sensitivities to model assumptions in GAMs

Modeling decisions are also necessarily made in the process of fitting GAMs and it is possible for these decisions to influence the outcome of an analysis. Within the context of our model, these decisions include the assumed error structure, the treatment of the uncertainty associated with the estimates of mortality, and the allowed shape of the smooths estimated in the GAMs. The following sensitivities address the implications of these modeling decisions on the outcome of our analysis.

Error structure

Impacts of assumptions about error structure were explored by assuming beta distributed data in the GAM and transforming the continuous total mortality rates to an exploitation rate ranging from 0 to 1. This transformation still resulted in mature population and temperature being the most important variables related to mortality, however the deviance explained decreased to 60% and 68% for mature and immature mortality, respectively, compared to 72% and 78% for the model presented in the main text. A potential shortcoming of the method presented in the main text is that predicted mortality could be less than zero. This was not the case in any of the model fittings, and would present a problem primarily if the model was extrapolated to data beyond the observed ranges. A potential fix to this issue is to log the response variable, but this resulted in unacceptable patterns in the residuals, so this model was not used.

Does incorporating the uncertainty in the estimates of mortality change model outcomes?

The models presented in the main text use the maximum likelihood estimates of mortality from the population dynamics models as response variables in the GAMs. However, each of those estimates of mortality have associated uncertainty estimated in the fitting process. One way of evaluating the impact of incorporating the estimated uncertainty from the population dynamics model into the GAM fitting process can be accomplished in 4 steps:

1. Invert the Hessian matrix produced from fitting the population dynamics model to calculate a covariance matrix describing the relationships between each of the estimates of mortality,
2. Simulate time series of the estimated mortality from a multivariate normal distribution with a mean of the point estimates of mortality deviations and the product of step 1 as the covariance matrix,
3. Refit the GAMs to these simulated mortality time series and record the deviance explained and p-values for each covariate,
4. Repeat these steps many times.

A similar methodology can be seen in Johnson et al. (2022). Each of these steps were taken in the R programming language. Inverting the Hessian was accomplished by using the function 'solve()'; simulating a time series of mortality deviations was accomplished using the function 'mvrnorm()'. The resulting simulated time series of mortality were very similar to the maximum likelihood estimates of mature and immature mortality, reflecting relatively precise estimates of mortality (Figure 46). The deviance explained across simulated time series were also similar to that produced with the MLE time-series of mortality. Temperature and mature population remained the most important variables in predicting mortality across GAMs fitted to simulated time-series of mortality (Figure 47). Given this outcome, the model presented in the main text does not consider the uncertainty associated with treating the estimates of mortality as 'data' in the fitting of the GAMs.

Shape of estimated relationships

The model presented in the main text constrains the number of knots available to the GAM to fit the data for each covariate to 3, but the shapes of GAM-estimated smooths are not constrained. This modeling choice was made because it is not immediately clear a priori what the shape of the smooths should be. For example, the relationship between immature population size and immature mortality could conceivably be linear positive (e.g. higher populations result in higher mortality due to intraspecific competition), linear negative (e.g. larger population sizes dilute the impact of external stressors like predation and fishery effects), dome-shaped (somewhat harder to interpret, but perhaps different processes are important at different population sizes), or monotonic in either direction (e.g. population size modulates external stressors to a point, after which other processes are more important).

The model in the main text is unconstrained with respect to the shape of the estimated relationships between mortality and covariates. However, the relationship between immature mortality and mature population size was markedly dome-shaped and a satisfying biological explanation for this shape is not immediately apparent. To explore the impacts of unconstrained estimation of this relationship, we refit the models using shape constrained additive models in the R package 'scam' (based on Pya and Wood, 2015). This allows the user to specify constraints on the shape of relationships between model variables (e.g. monotonically increasing or decreasing). We refit our model for immature mortality with the assumption that the relationship between immature mortality and mature population size can only be monotonically increasing. The rest of the covariates were specified as linear predictors, except temperature, which remained non-linear. Given the importance of temperature in the hypotheses we present, we were particularly interested to understand how the assumptions about the shape of other significant covariates influence the estimated relationship between temperature and mortality.

Temperature and mature population size were still significant covariates within the shape constrained additive model and immature population size became significant (Table 5). The estimated relationship between mortality and temperature was still strongly positive, but became more linear with the shape constraints imposed on mature population size (Figure 48). The relationship between immature population size and immature mortality was negative (i.e. all other things considered, more immature crab were associated with lower mortality). While the immature population relationship is potentially interesting, the most important outcome of this exercise is that temperature still returned positive relationship with estimated mortality. Using temperature as the only covariate in an unconstrained GAM explained 37% and 38% of the deviance in immature and mature mortality (not shown). All of these points suggest temperature is a key covariate in the estimated mortality dynamics for snow crab in the eastern Bering Sea.

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

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	-0.0325	0.1850	-0.1757	0.8623
disease	0.0236	0.0162	1.4561	0.1607
imm_pop	-0.0374	0.0174	-2.1454	0.0442
predation	0.0002	0.0169	0.0137	0.9892
bycatch	-0.0003	0.0147	-0.0194	0.9847
cannibalism	-0.0090	0.0153	-0.5887	0.5626
B. smooth terms	edf	Ref.df	F-value	p-value
s(temperature)	1.0005	1.0009	6.7766	0.0167
s(mat_pop)	1.7664	2.0738	3.5320	0.0466

Table 5: GAM output for trimmed model predicting immature mortality. Deviance explained = 59.53 %

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 modeled fraction of 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, $kCal_{s,y}^{snow}$, 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:

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

Caloric requirements increased sharply in 2018 and to explore potential impacts of this increase, we analyzed the weight at size data available (Figure 49). A GAM was 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 GAM explained 97.4% of the deviance in the weights of snow crab and all covariates were significant (Table 6).

In general, higher temperatures were associated with higher weight at size (Figure 50). 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

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

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. Previous studies looking at the impacts of starvation on the weight at size of snow crab (e.g. Hardy et al., 2000) reported small changes in weight at size (roughly 2.6% of weight lost over 5 months), but larger changes in the weight of the hepatopancreas. However, there are some key differences between these studies and the field observations we report. First, the maximum observed mortality was 20% in the starvation studies; the mortality levels estimated in the Bering Sea exceeded 90% in some years. Second, the starvation experiments were in laboratory environments where no foraging occurred. Seventy crab were confined in containers measuring 122 x 183 x 40 cm in Hardy et al. (2000), which greatly restricted movement and would presumably impact caloric expenditure and the initiation of catabolism of muscle tissue.

A word on methods

Attribution of changes in population processes in ecology is a difficult problem, particularly for wild populations that are difficult to directly observe and impossible to experiment on in situ. There are a large range of methodologies that claim to identify causality in observational data (structural equation modeling, empirical dynamic modeling, etc.). Some of the difficulties in determining causality in ecological time series are related to the generally short time series that are available, non-linear dynamics, and departures of populations or covariates into unexplored parameter space. These issues can present issues for any modeling framework and we have tried to address these the the best of our ability with the models used here. The use of p-values has been (rightfully) criticized in the literature and the explanatory power of our models are likely overstated. Ultimately, the numerous sensitivities and simulation tests performed here were undertaken to try to understand if a suite of covariates appear to be important under different modeling decisions and considerations of uncertainty, in spite of the potential short-comings of the data available and models selected. Temperature and population density proved to be these covariates.

Frequently asked questions

Are you sure the collapse wasn't a result of cod predation?

The predation index (i.e. the crab consumed by cod divided by the crab available; Figure 19) was near the time series average during the collapse in 2018 and 2019. If predation were a strong driver of the mortality during the collapse, it is difficult to explain why estimated mortality was not high when the predation indices were much higher in the late 1990s and mid-2010s. Furthermore, the distribution of the cod population during 2018 and 2019 extended much farther north, beyond the portion of the snow crab population that is included in this analysis. Movement north can happen in particularly warm years and would serve to reduce the relative predation pressure on the portion of the population of crab in this analysis because the cod that moved north would be consuming crab outside of our study area. Finally, a large fraction of the missing crab from the recent collapse were not of the sizes typically eaten by cod (Figure 51).

Although our predation indices incorporate the best available information about cod diet and abundance, these indices are snapshots taken during the summer survey. It is possible that the consumption of crab was different in other times of the year and, if this were true, knowing how predation changed throughout the year could alter our results.

Are you sure the collapse wasn't a result of trawling?

The bycatch index steadily declined since the beginning of our study, with the relative impact of trawling in 2018 and 2019 below the historical average (Figure 19). It is difficult to reconcile the idea that trawling could have contributed to the collapse with the relatively low mortality rates estimated during the periods when the bycatch index was many times higher in the 1990s. Furthermore, if trawling is a large source of mortality for snow crab, it is difficult to understand how the largest pseudocohort ever observed could have established and survived for ~8 years on the Bering Sea shelf, during which the trawling pressure was relatively consistent.

However, not all of the mortality associated with non-directed fleets is observed. The index used in this analysis is a reliable indicator of the trend in bycatch mortality provided the ratio of observed to non-observed mortality is consistent over time. If this is not the case, that could change the outcome of our analysis, but there is no clear methodology for determining that ratio.

What do crab eat? If they starved, did there appear to be large declines in their prey base?

Snow crab have a wide-ranging diet of bivalves, polychaetes, crustaceans, and gastropods in the northern Bering Sea (Kolts et al., 2013). They appear to be a generalist, consuming whatever they can capture and crush with their claws. Kolts et al. (2013) reported that most prey items were consumed in proportion to their estimated abundances, except polychaetes, which seemed to be preferentially selected. The prey items are relatively poorly sampled in the Bering Sea survey, so time series of prey quantity are unreliable.

Even if there were abundant forage for snow crab in warm years, the metabolic trade-off between the energy required to obtain, handle, and digest their prey and the energy derived from prey would need to be considered when trying to understand if metabolic demands could be met. This would be a useful area of further research, particularly if reliable time series of benthic forage could be started and maintained.

If it was a large mortality event, did you see large numbers of empty carapaces in the survey?

Hundreds of millions of carapaces are discarded by molting crab each year even when there are no mortality events and these are rarely seen in the survey nets. So, even with a massive mortality event, one might not expect to see the carapaces remaining from the event. Why the carapaces are not seen in the survey nets is not completely clear, but potential hypotheses include relatively fast disintegration on the sea floor or poor selectivity by the survey gear. Discarded carapaces may sit flat on the bottom and be passed over by the net.

Were that many crab really in the eastern Bering Sea to begin with? Was the 'collapse' an artifact of some survey error?

The NMFS summer survey was designed to estimate crab abundances (REFERENCE). Snow crab are widely distributed on the shelf and consequently well sampled by the survey. There are 375 survey locations in the NMFS eastern Bering Sea trawl survey, of which 349 are on a 40 square nautical mile grid. The remaining stations are in high density sampling areas around islands in the Bering Sea implemented to better estimate crab abundances around those islands. On average, snow crab are observed at 233 of the 349 survey stations on the standard grid. In 2018 a large number of stations returned estimated high densities of crab (see Fig. 1 of main text), which means that the large estimates of abundance in 2018 were not driven by one or two

large survey tows. Further, the recent survey methodology has been repeatedly verified as a useful tool for estimating crab abundance (see Somerton et al., 2013, for example).

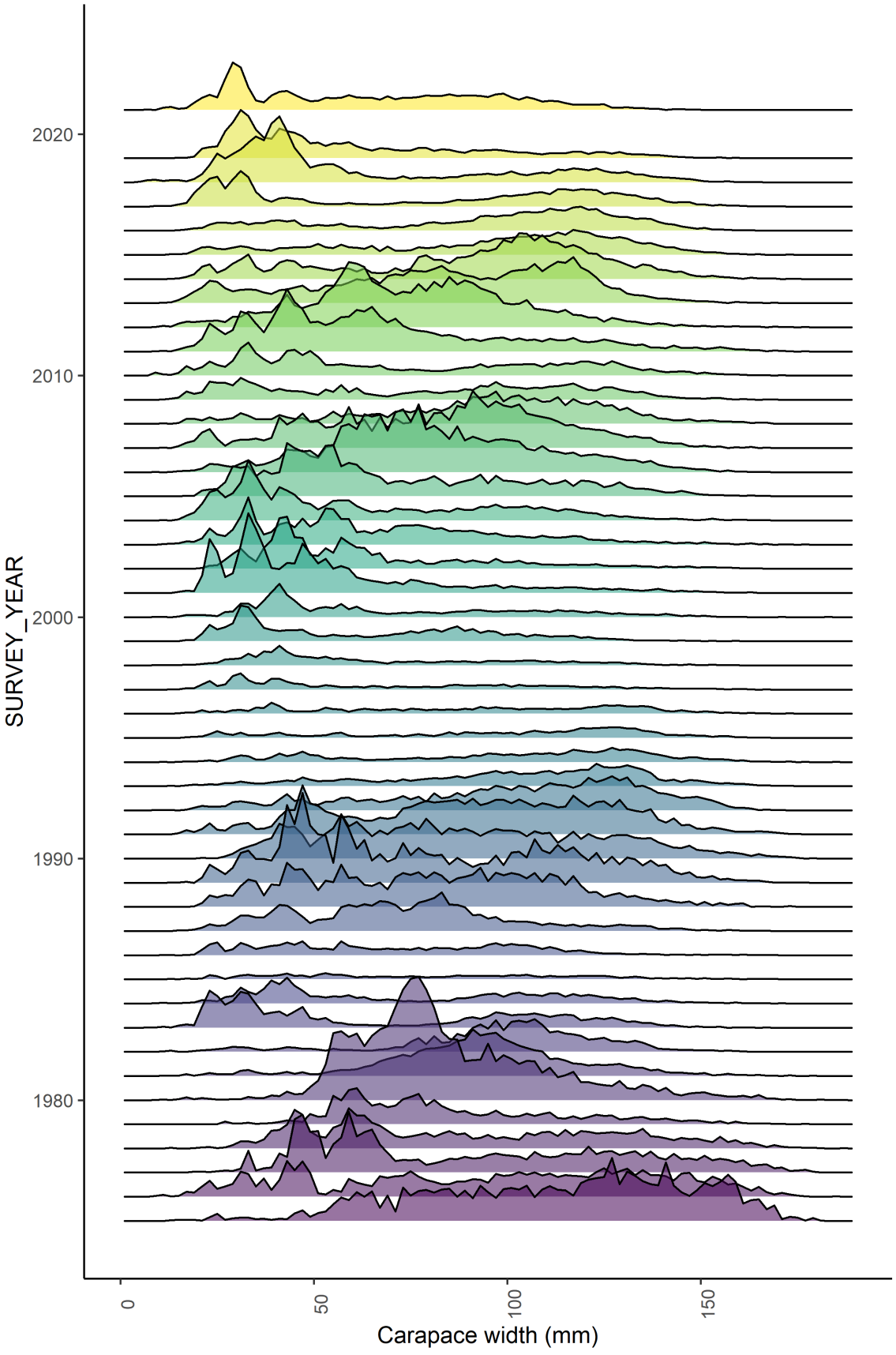

Figure 1: Observed abundance by carapace width of Tanner crab in the NMFS summer survey.

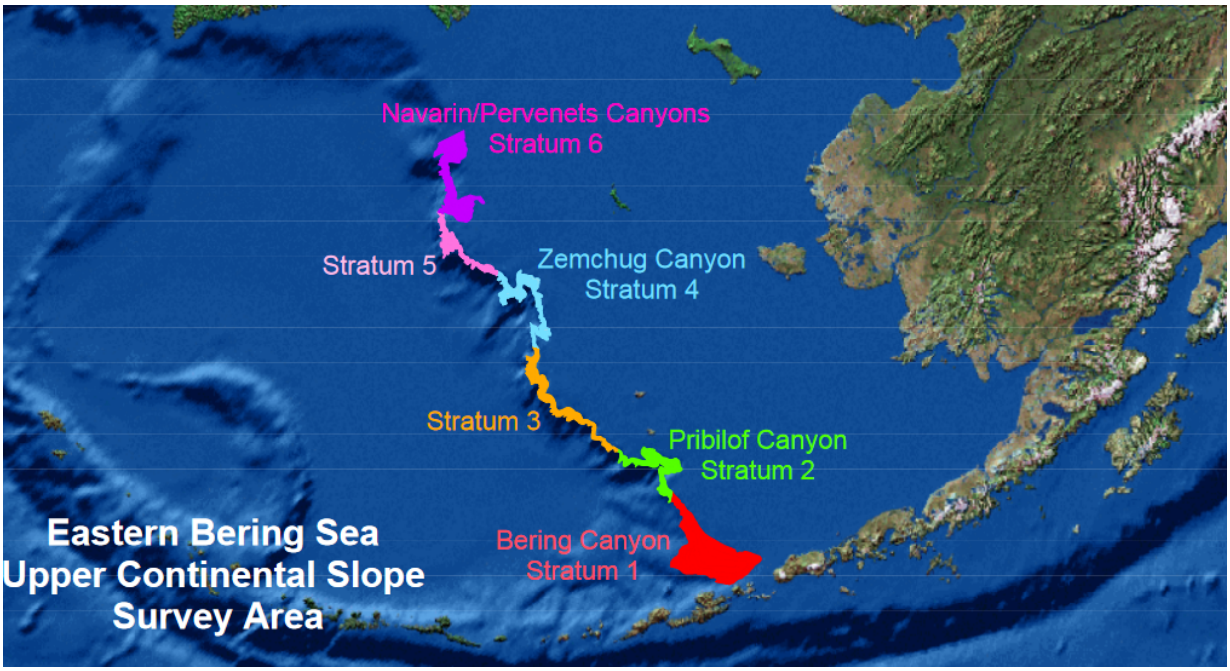

Figure 2: Map of eastern Bering Sea slope habitat (colors). Reproduced from HERE.

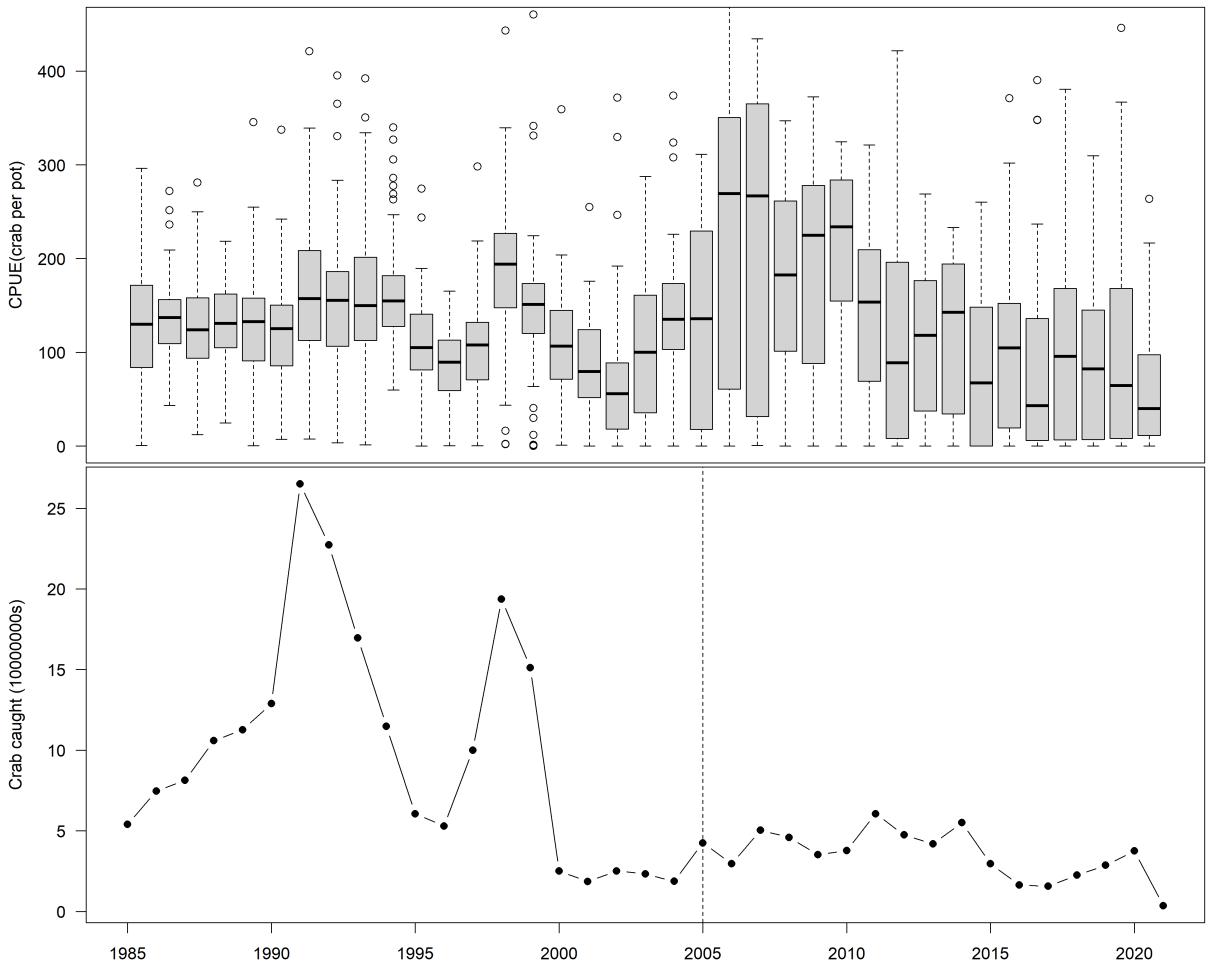

Figure 3: Fishery cpue (top; black lines are median, grey box represents 25-75th quantiles, circles are outliers) and number of crab caught in the directed snow crab fishery (bottom). Vertical dashed line represents the introduction of individual transferrable quota management.

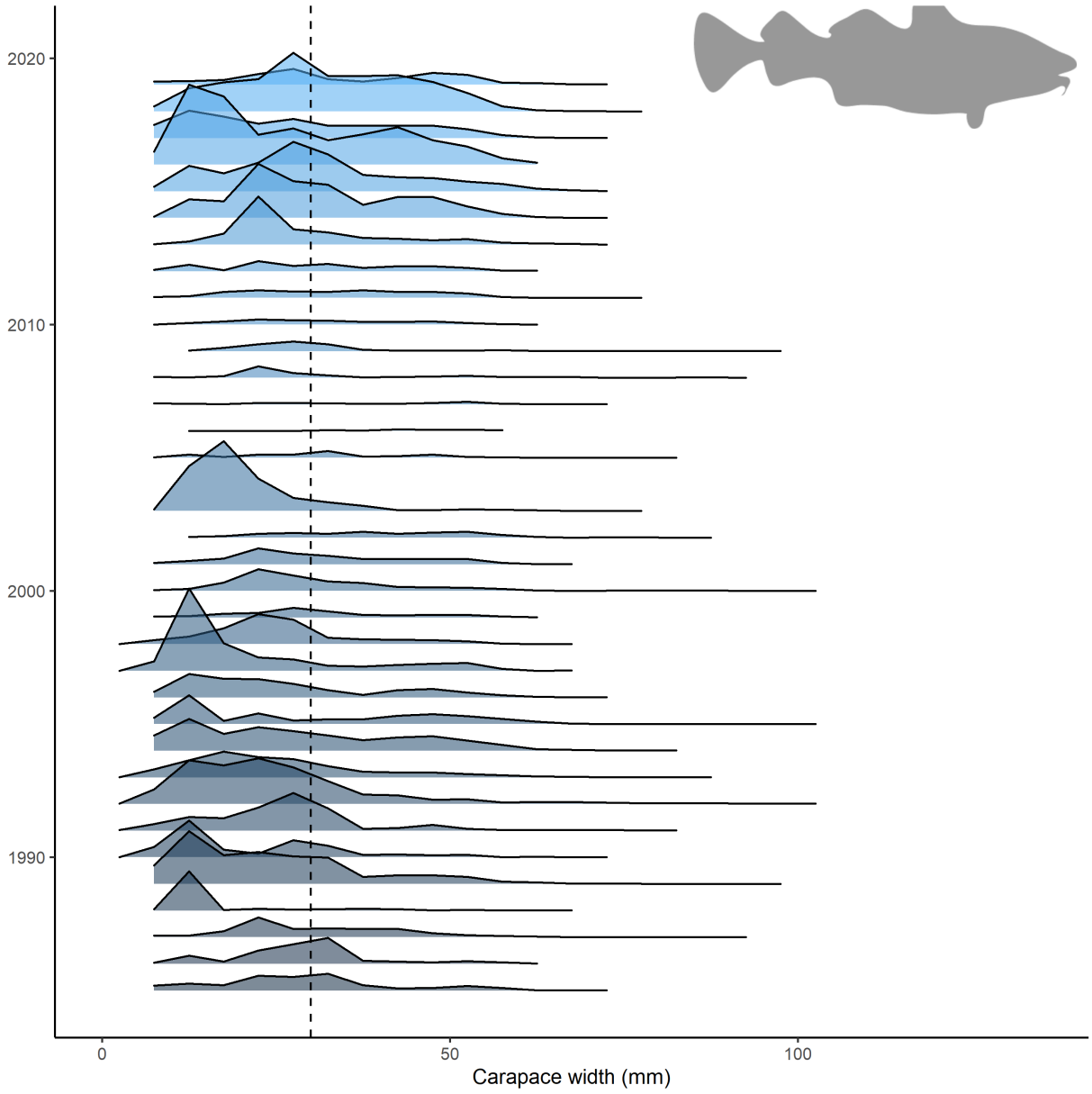

Figure 4: Consumption of crab by Pacific cod at size over time. Dashed line represents the size at which crab enter the population dynamics model presented in the text.

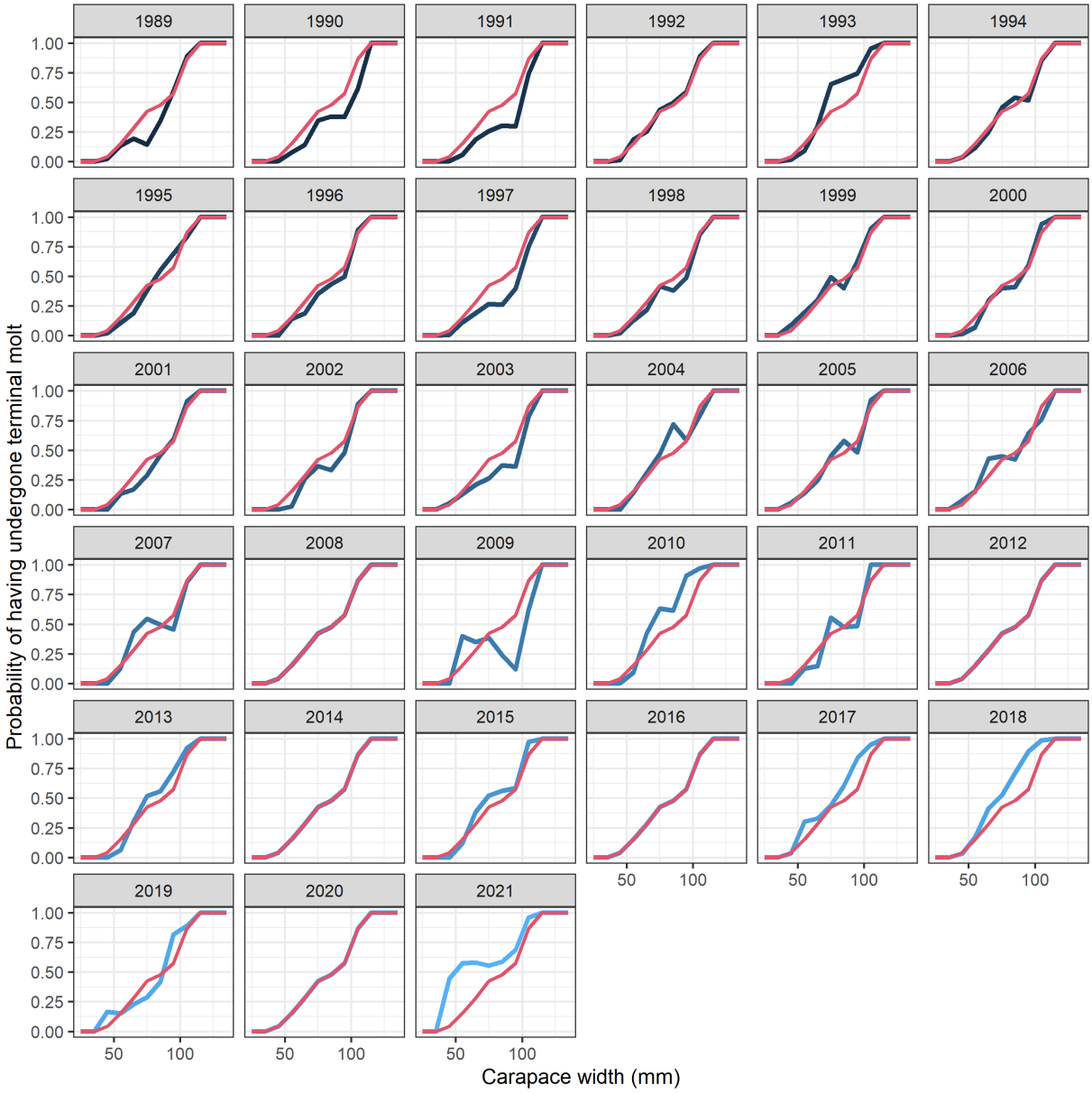

Figure 5: Observed proportion of mature new shell crab in the NMFS summer survey. Red line represents the median over years and the blue lines are the observed data. Chela height data were not collected in years without a blue line. These data are used to separate the numbers at size into mature and immature states for the input data to the population dynamics model.

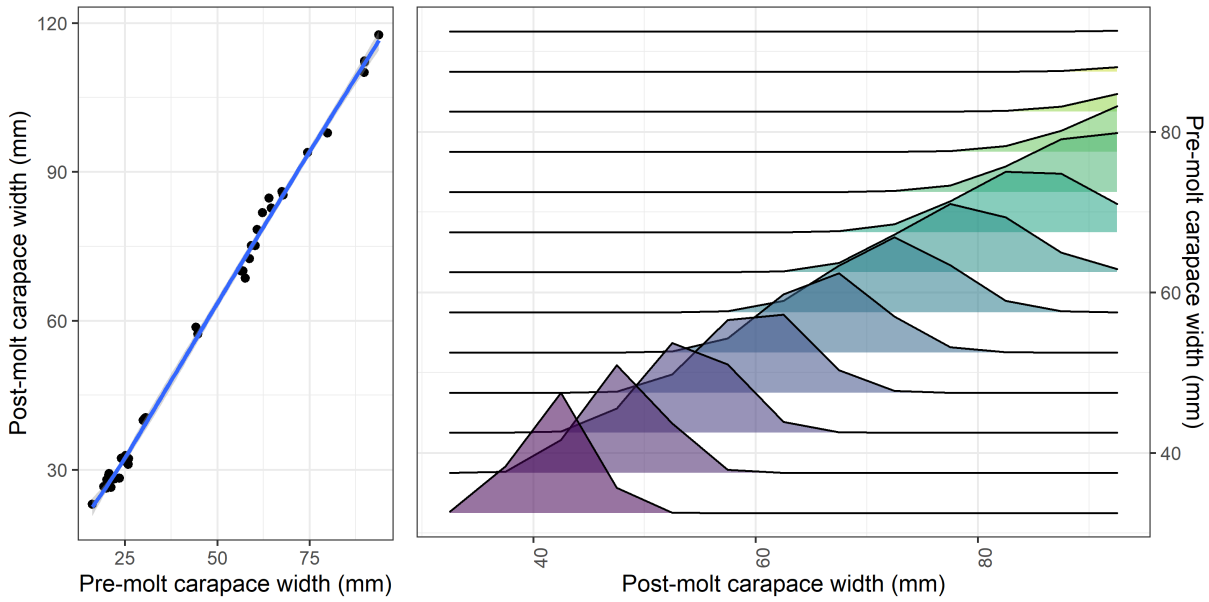

Figure 6: Empirical relationship between pre- and post-molt size (left) derived from crab captured in the wild pre-molt and observed to molt in the lab. Calculated size-transition matrix used in the population dynamics model (right).

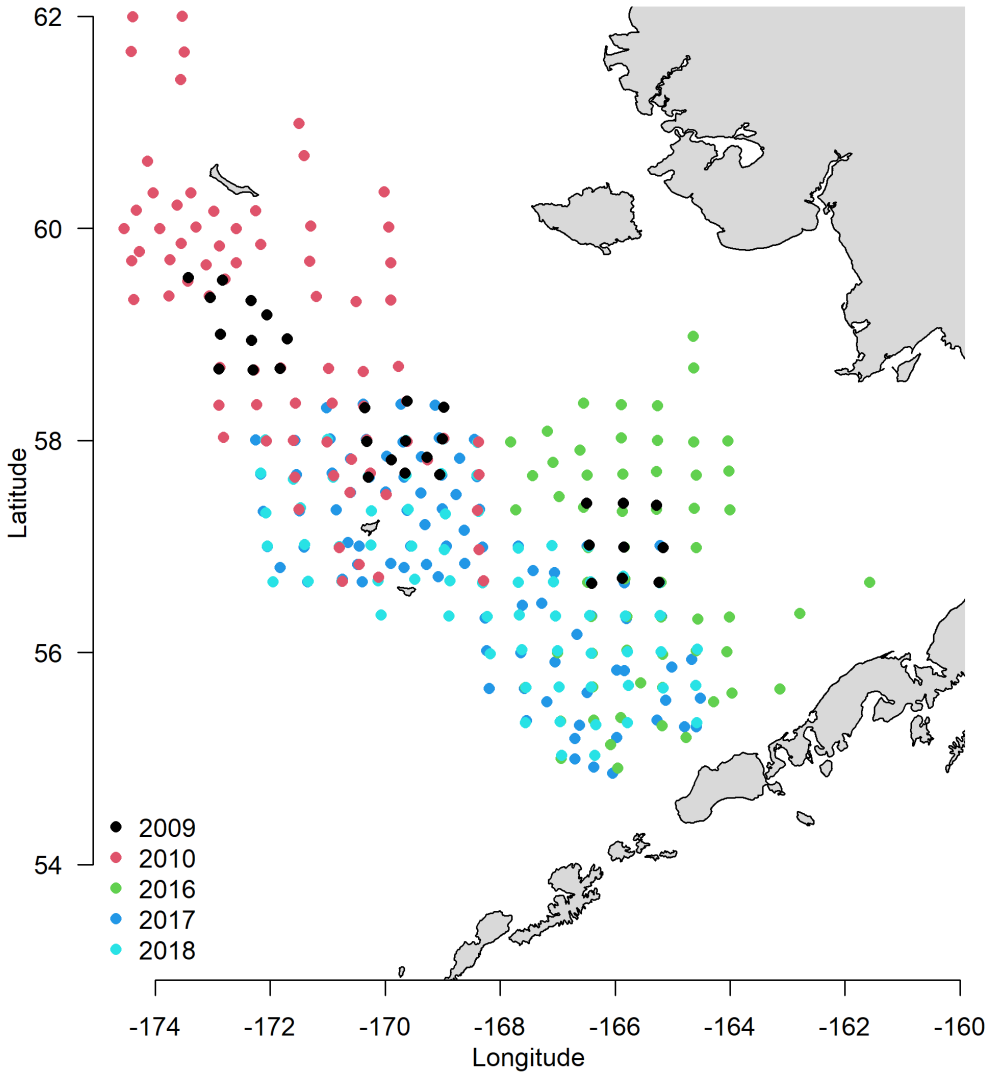

Figure 7: Locations of the BSFRF experimental trawls to evaluate the capture efficiency of the NMFS summer trawl survey for snow crab in the eastern Bering Sea.

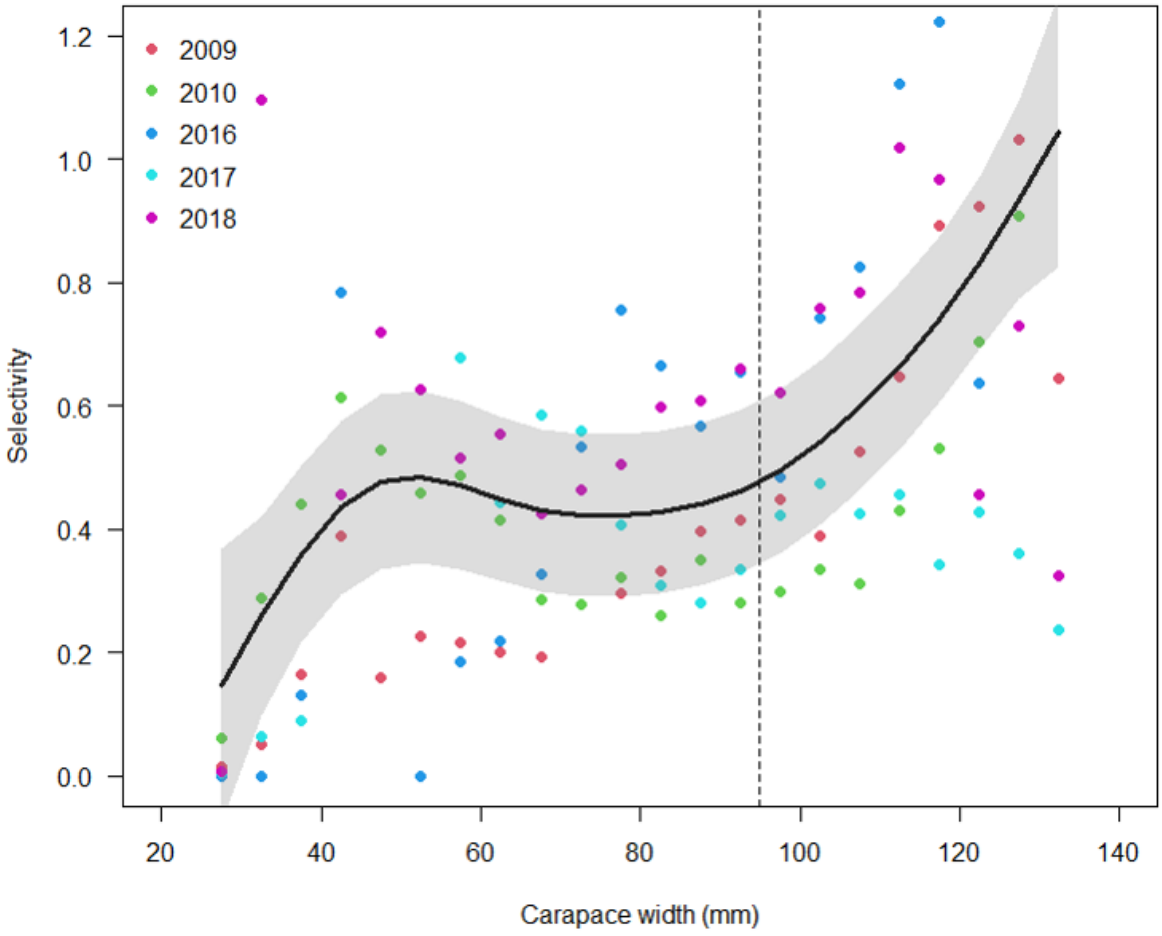

Figure 8: Inferred selectivity from the BSFRF experimental trawls.

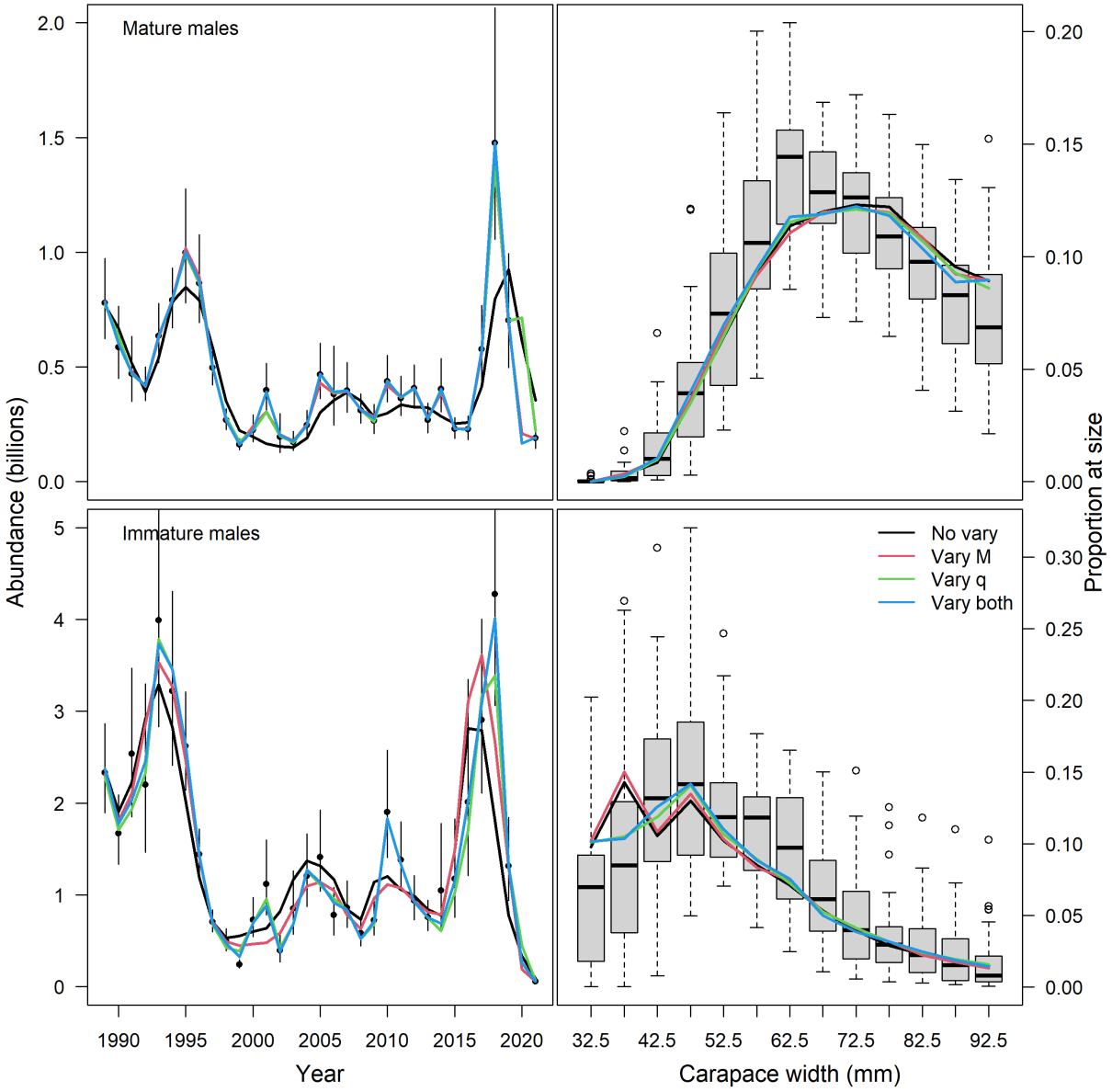

Figure 9: Fits of models with increasing complexity in mortality and catchability. Index of abundances are on the left with observations in black dots with 95% confidence intervals; colored lines are model fits. Size composition data are at the right with observations in box plots (aggregated over year; black lines are median, grey box represents 25-75th quantiles, circles are outliers) and colored lines are model fits.

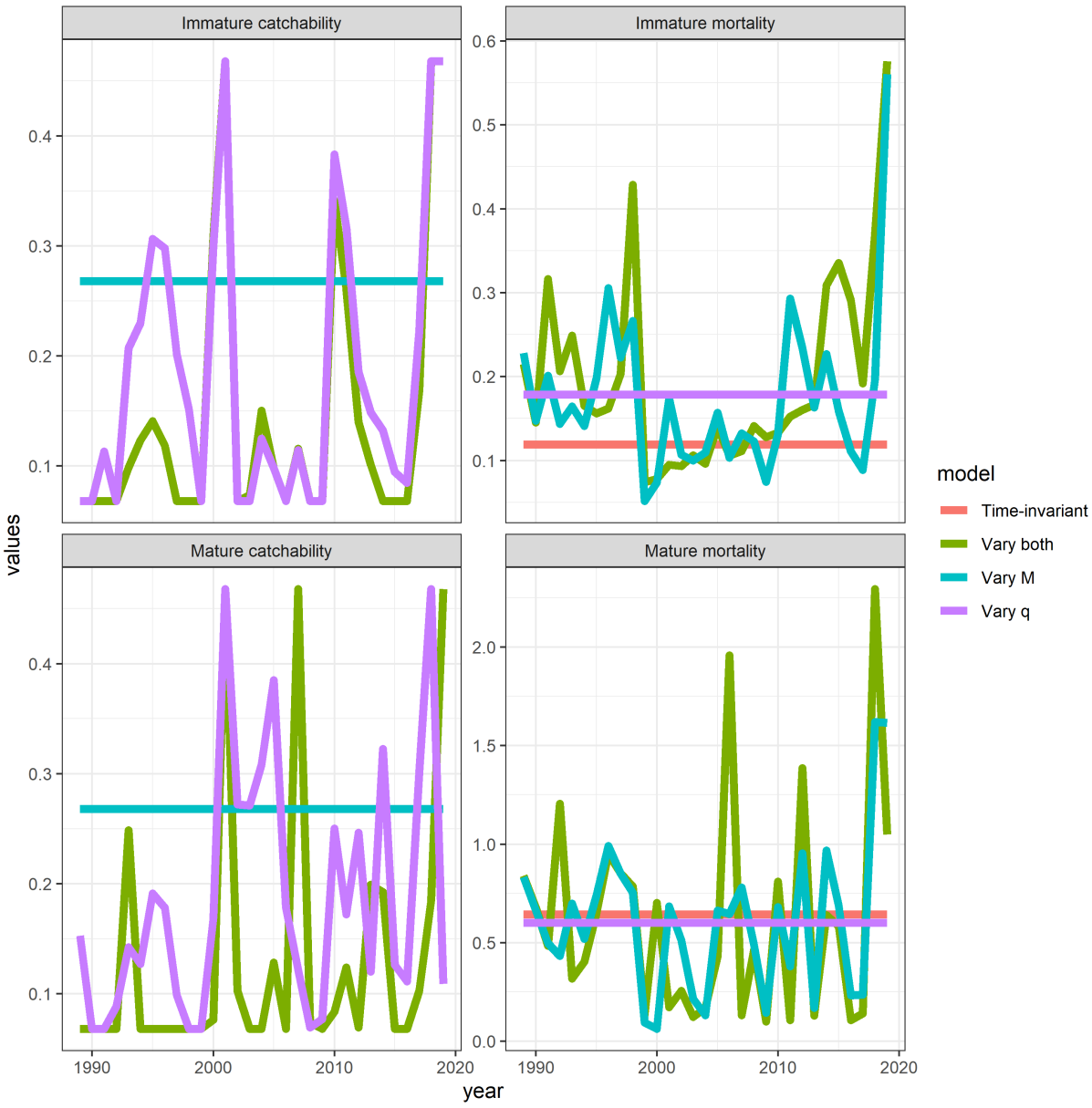

Figure 10: Estimated processes from model with increasingly complex time-variation in mortality and catchability.

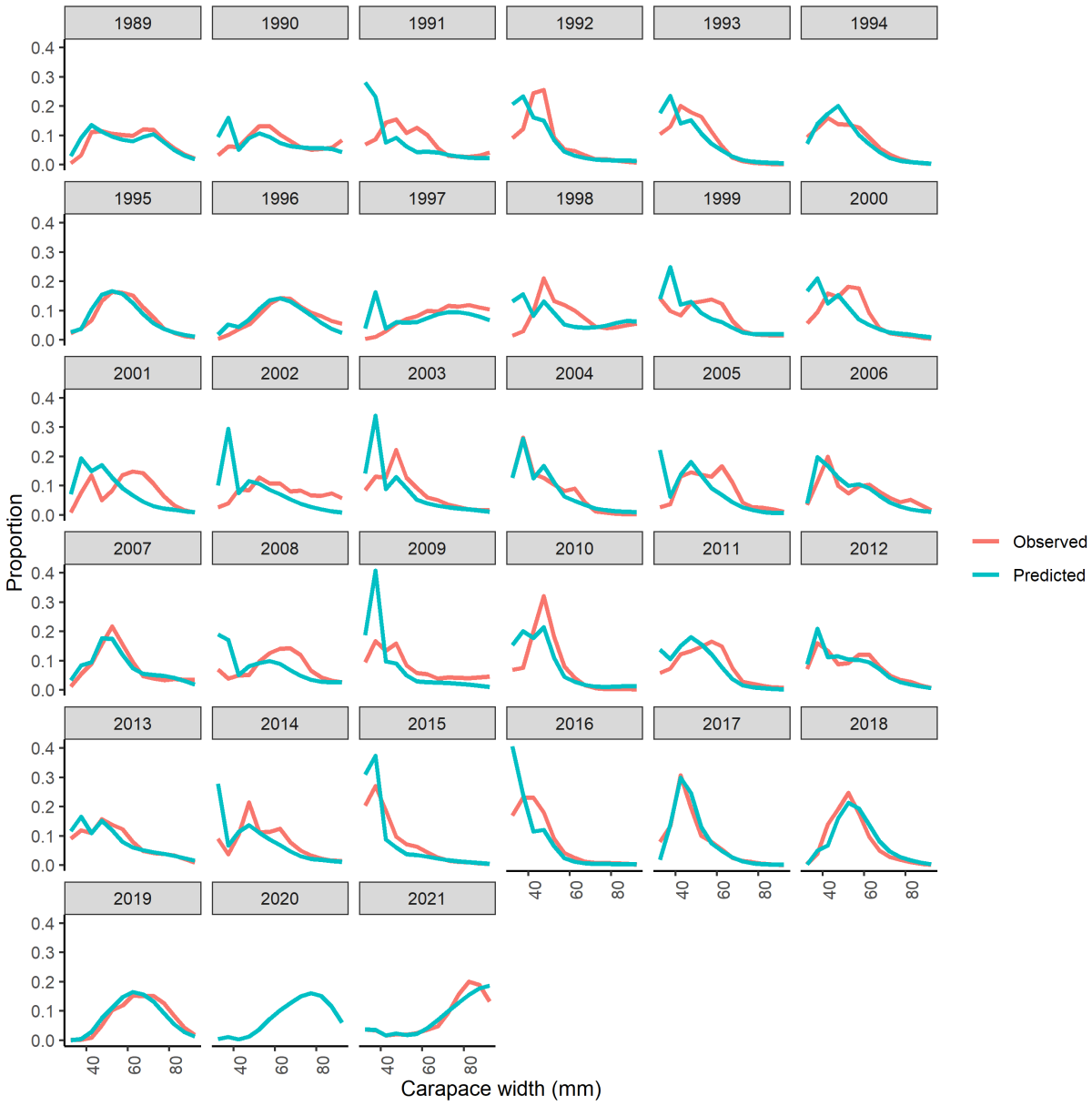

Figure 11: Fits for individual years to immature size composition data from a model in which mortality varied over time.

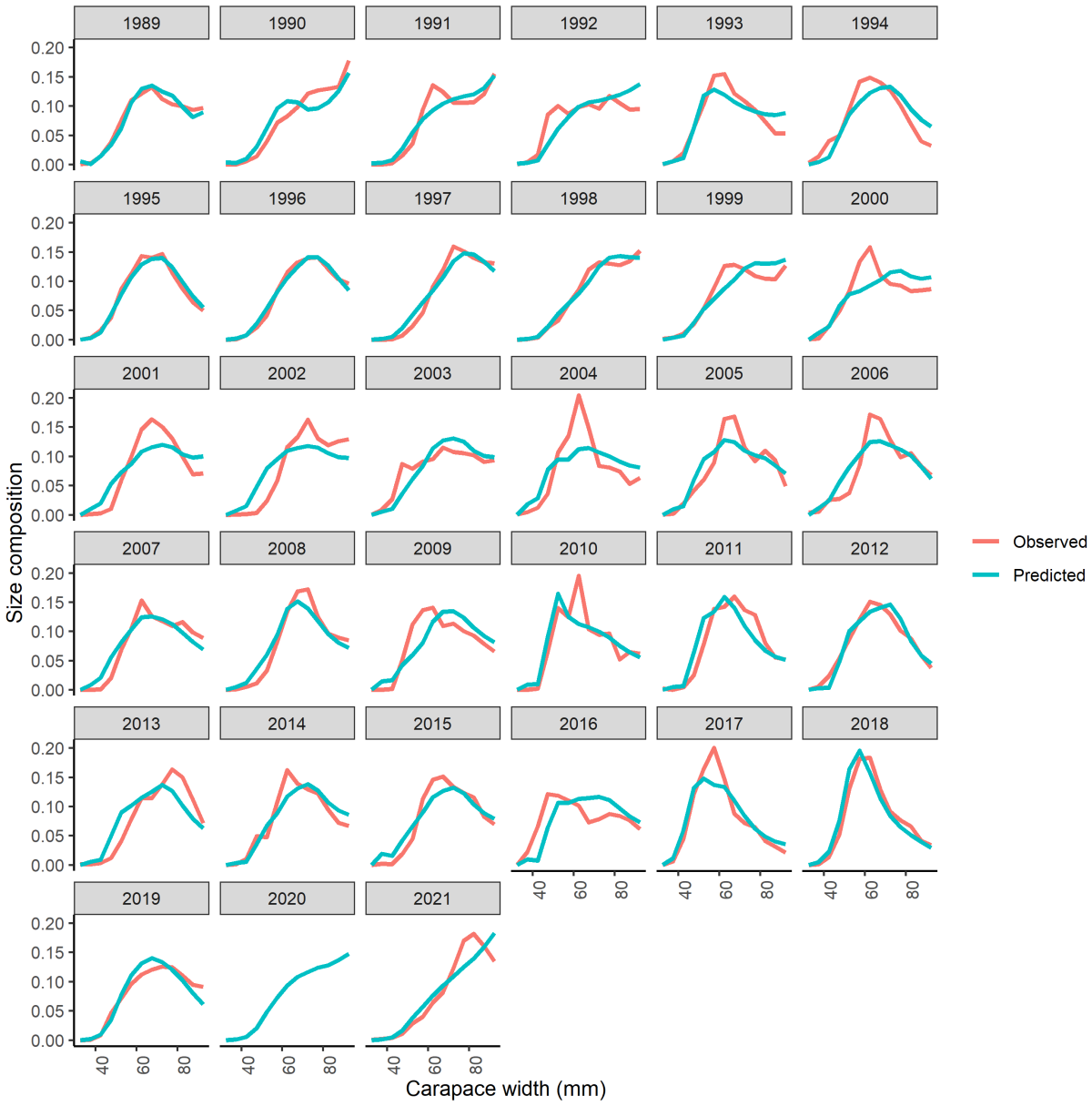

Figure 12: Fits for individual years to mature size composition data from a model in which mortality varied over time.

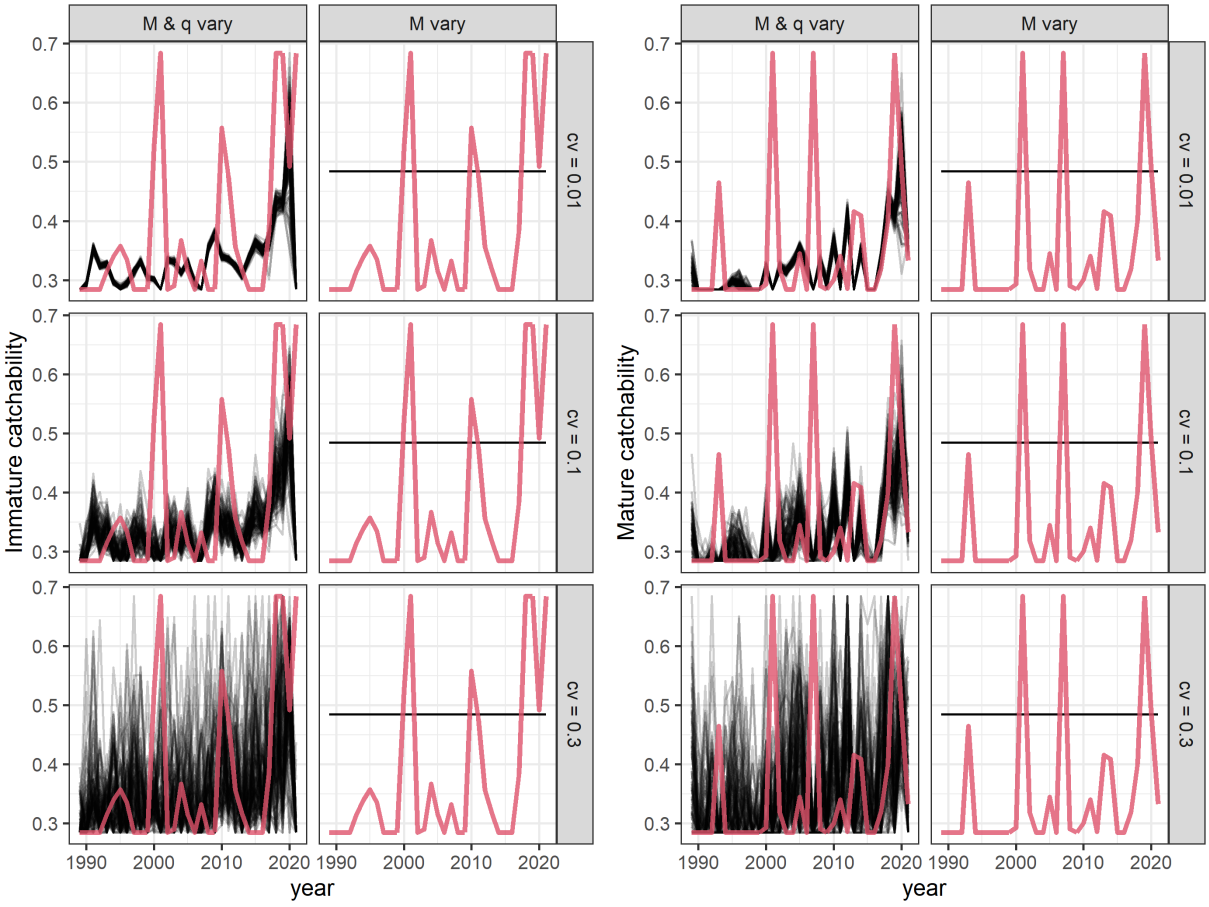

Figure 13: Estimates of catchability by maturity state (black lines) compared to the underlying values (red line) from simulations testing the estimation ability of the population dynamics models.

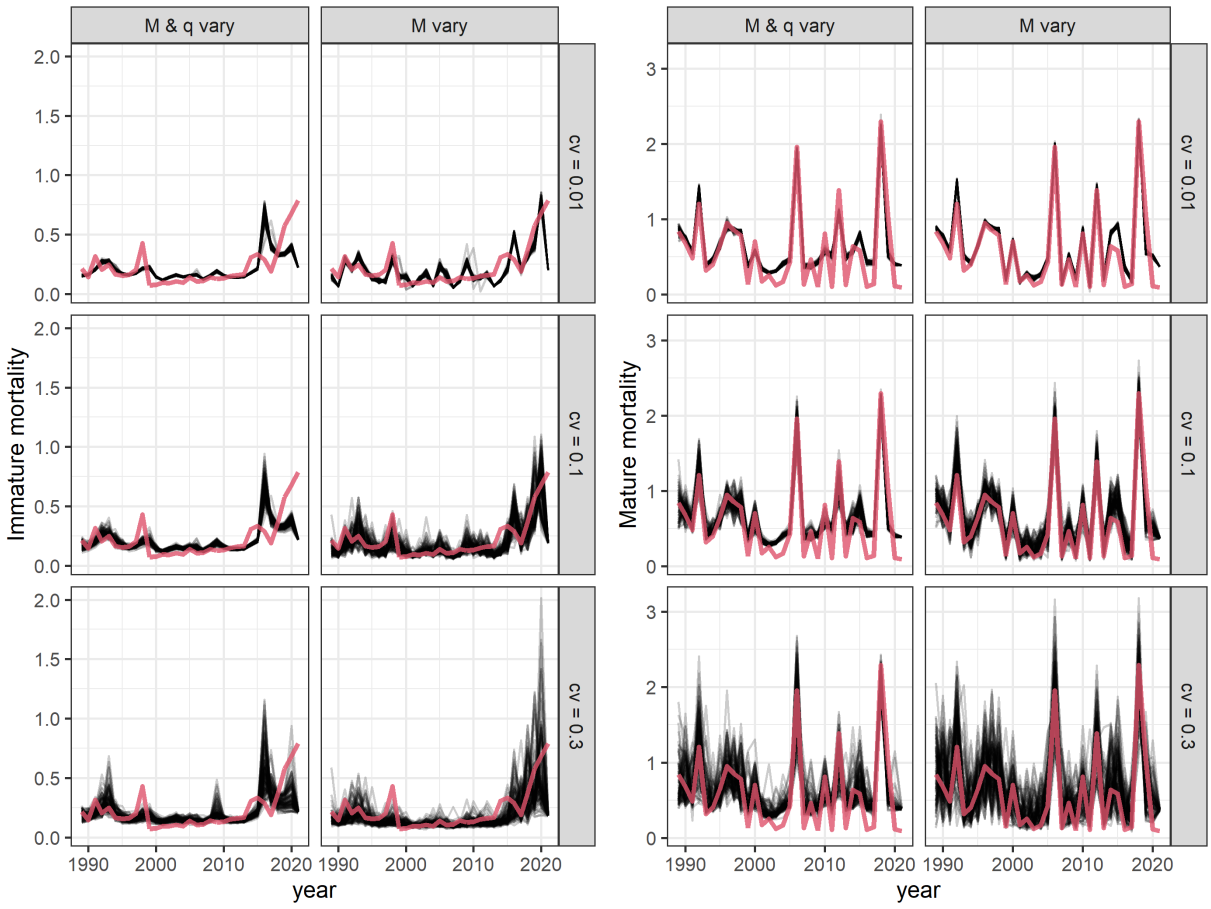

Figure 14: Estimates of mortality by maturity state (black lines) compared to the underlying values (red line) from simulations testing the estimation ability of the population dynamics models.

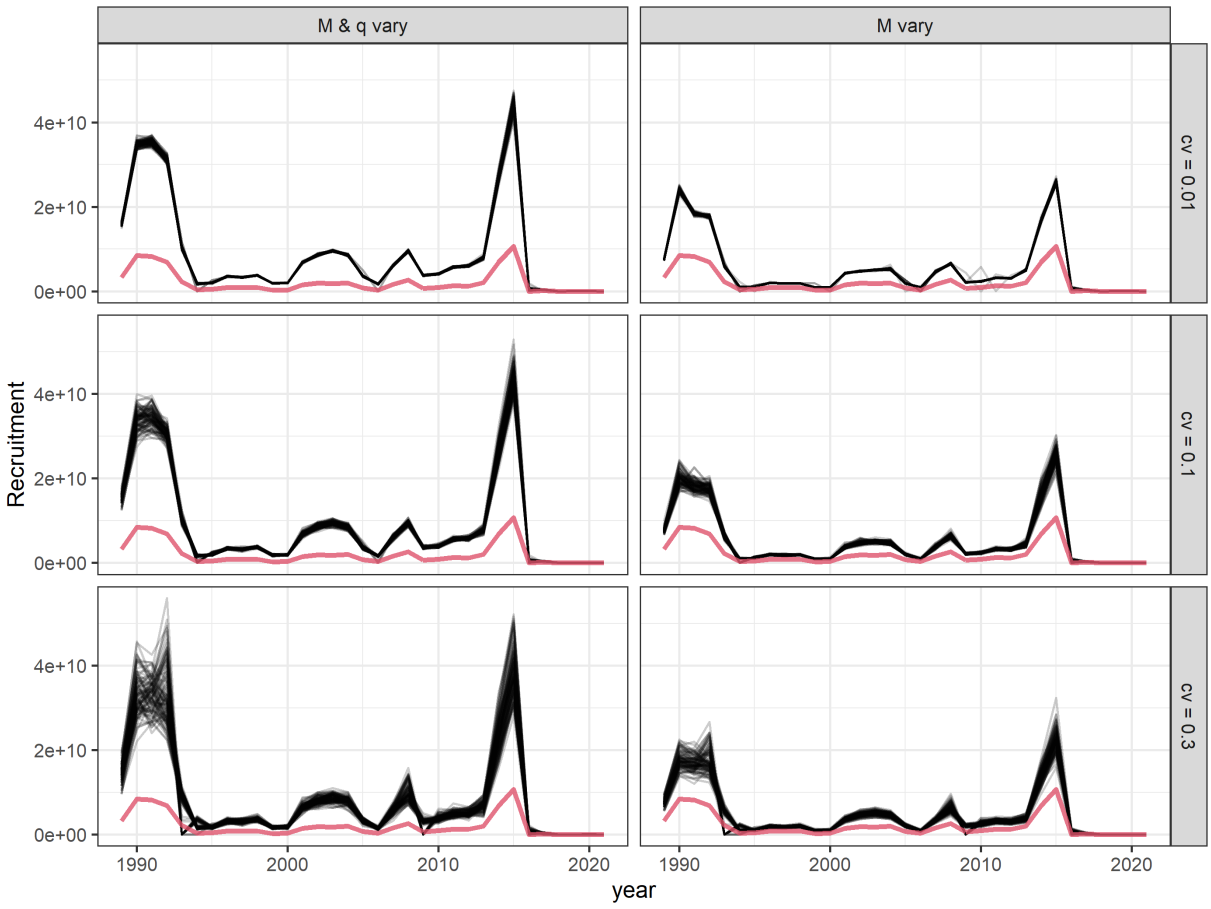

Figure 15: Estimates of recruitment (black lines) compared to the underlying values (red line) from simulations testing the estimation ability of the population dynamics models.

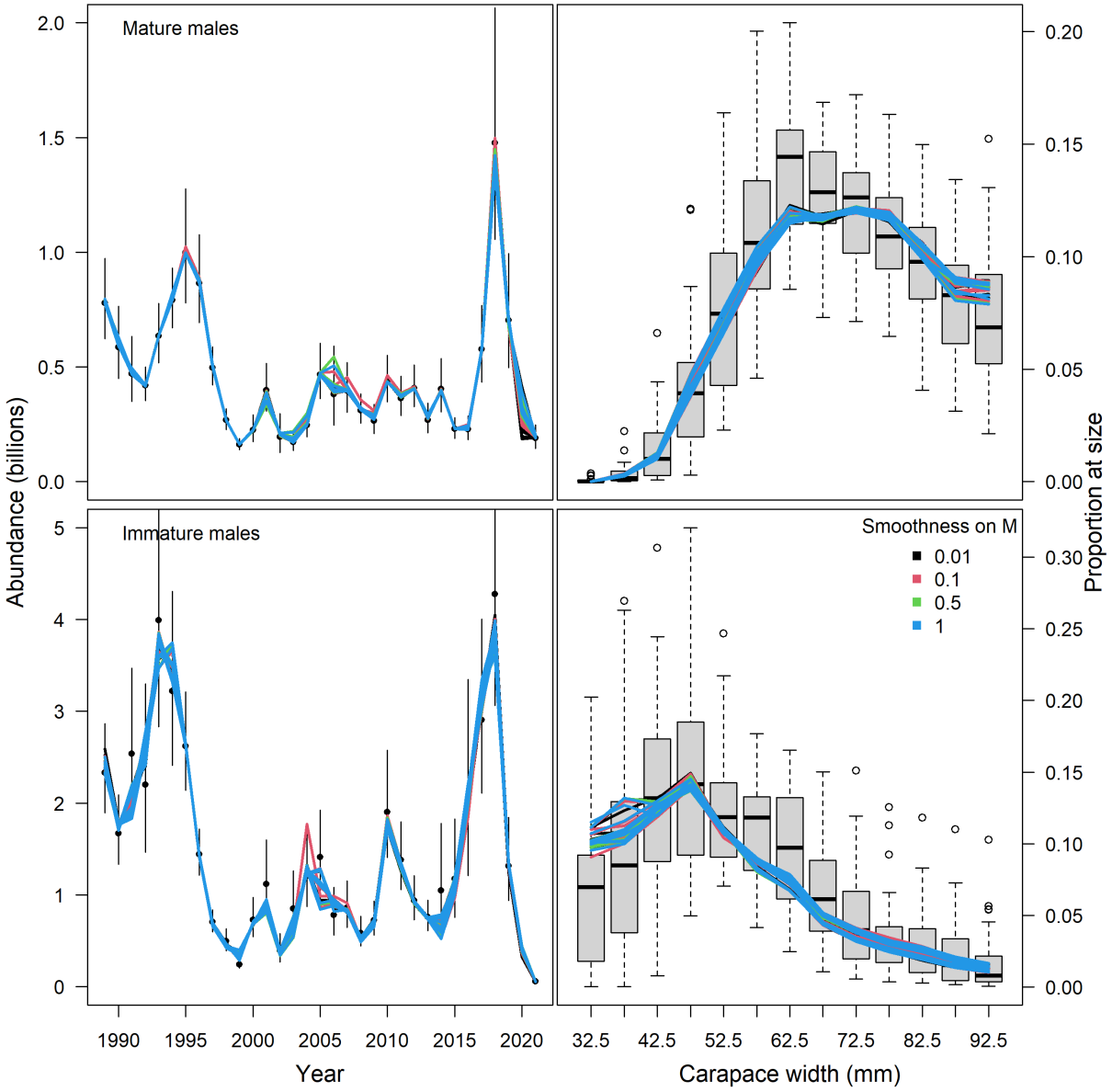

Figure 16: Model fits from sensitivity tests. Indices of abundances are on the left with observations in black dots with 95% confidence intervals; colored lines are model fits. Size composition data are at the right with observations in box plots (aggregated over year; black lines are median, grey box represents 25-75th quantiles, circles are outliers) and colored lines are model fits.

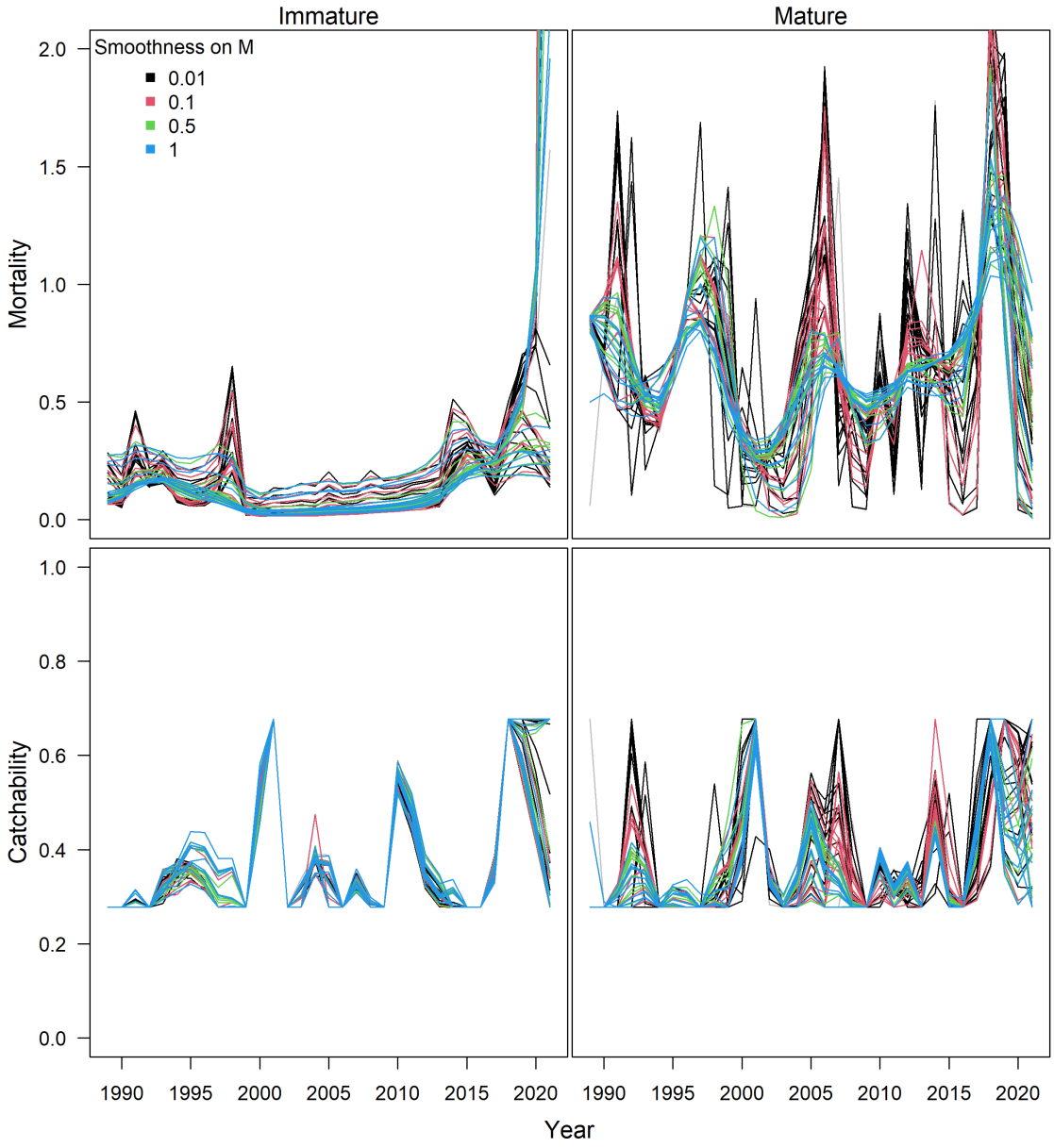

Figure 17: Estimates of mortality and catchability by maturity state over sensitivity runs. Lines are colored based on the smoothness penalty on mortality.

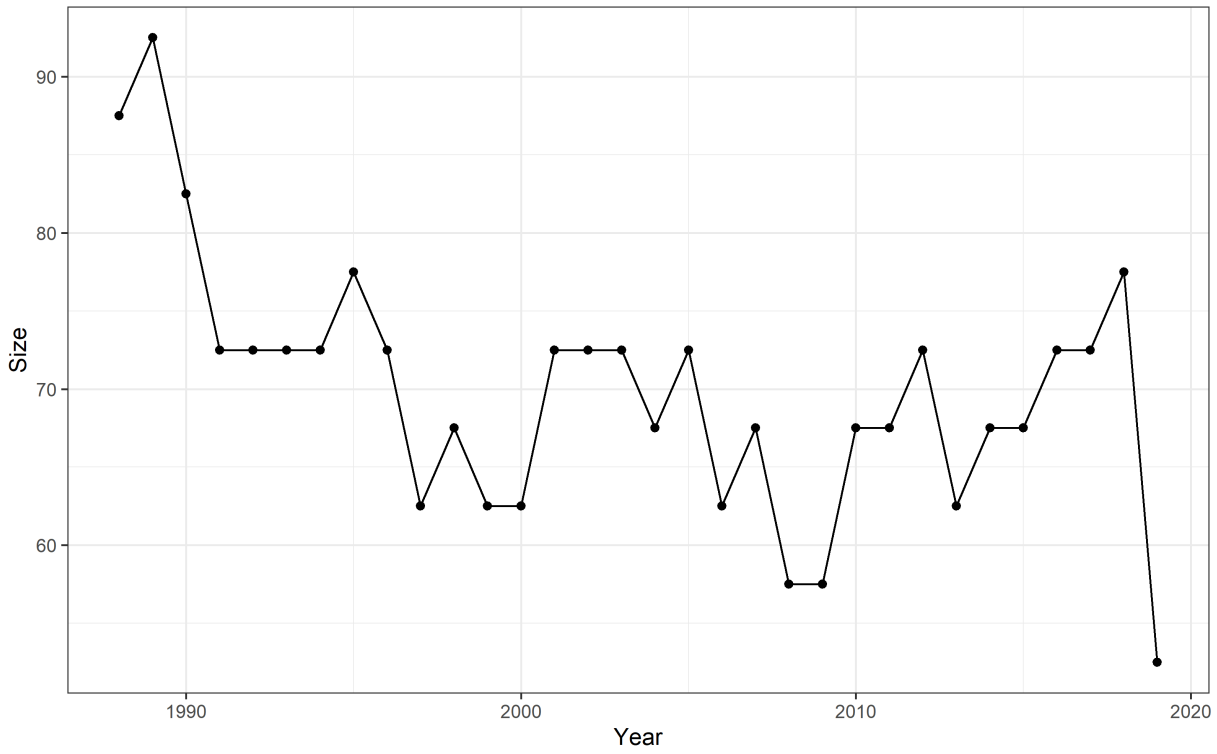

Figure 18: Size at which half of the crab in the population are mature over time (note, this is not the probability of undergoing terminal molt, rather the proportion of the number of mature vs. immature crab at size in the population).

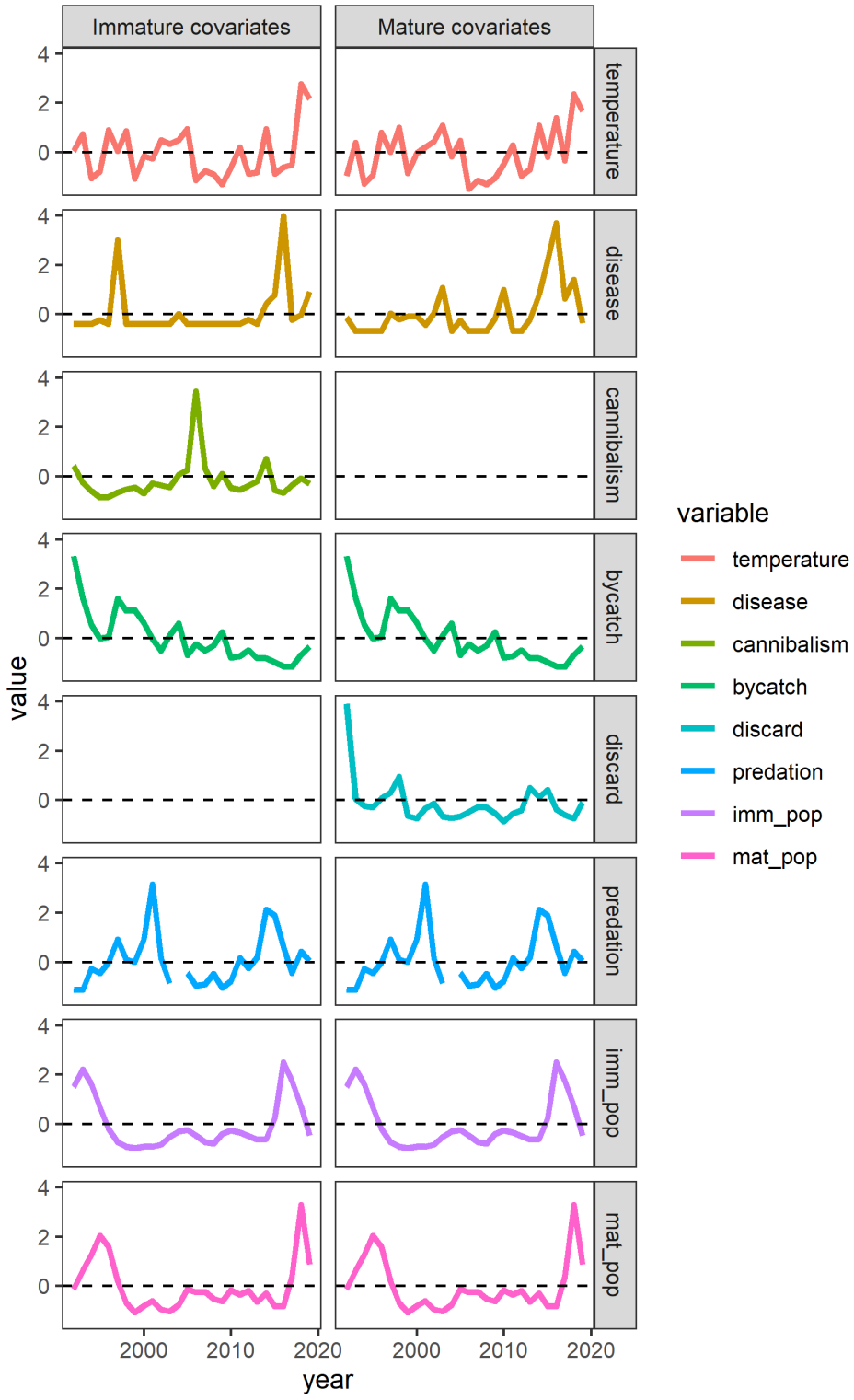

Figure 19: Calculated covariates incorporated into GAMs to relate stressors to estimated mortality. Two covariates (discard and predation) are only relevant for one maturity state based on the critical role size plays in the process (i.e. discards are primarily relatively large crab and predation is primarily smaller crab).

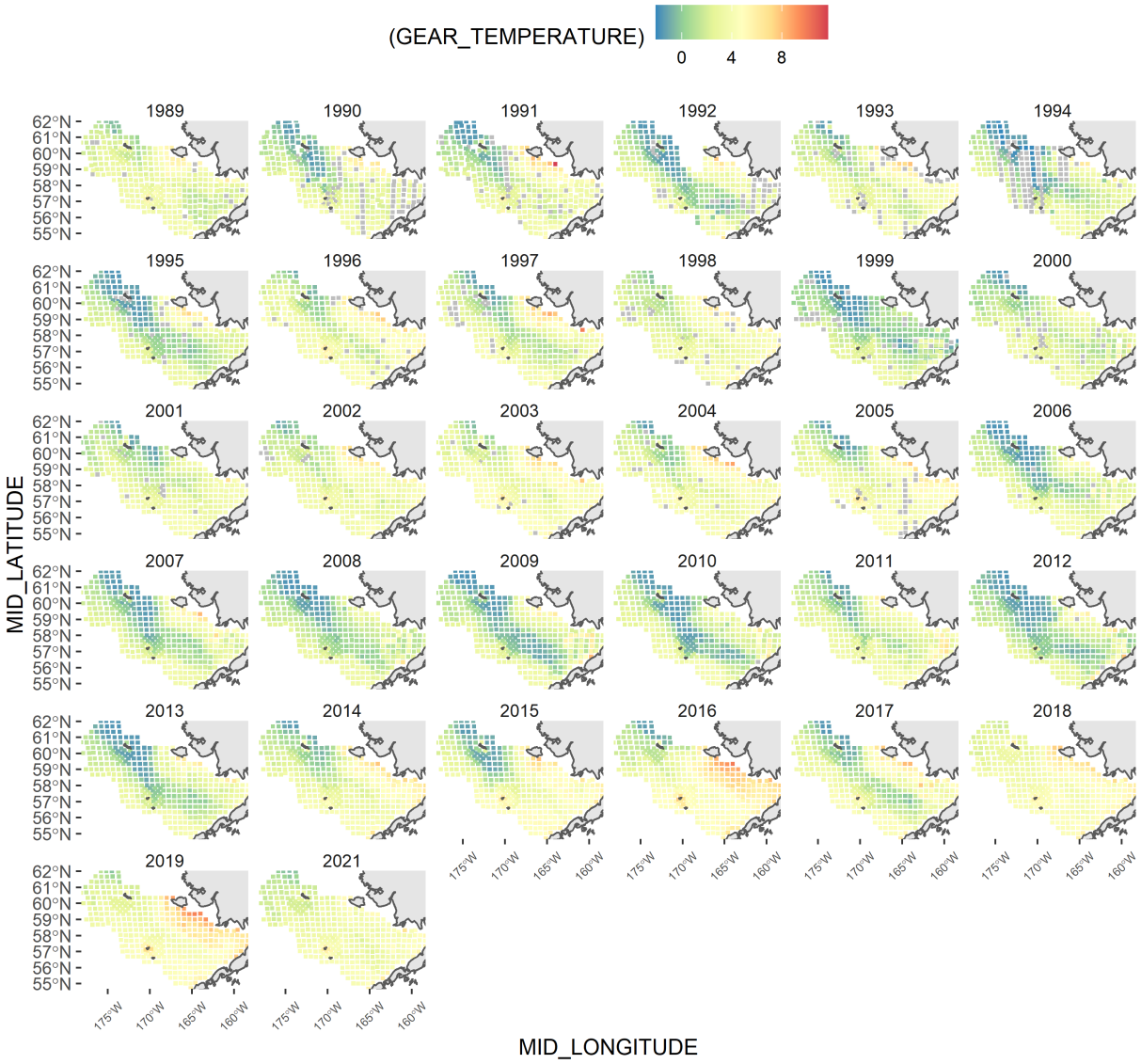

Figure 20: Observed bottom temperature at the time of the NMFS summer survey. Less than 2 degrees C represents the cold pool, seen in green and blue here.

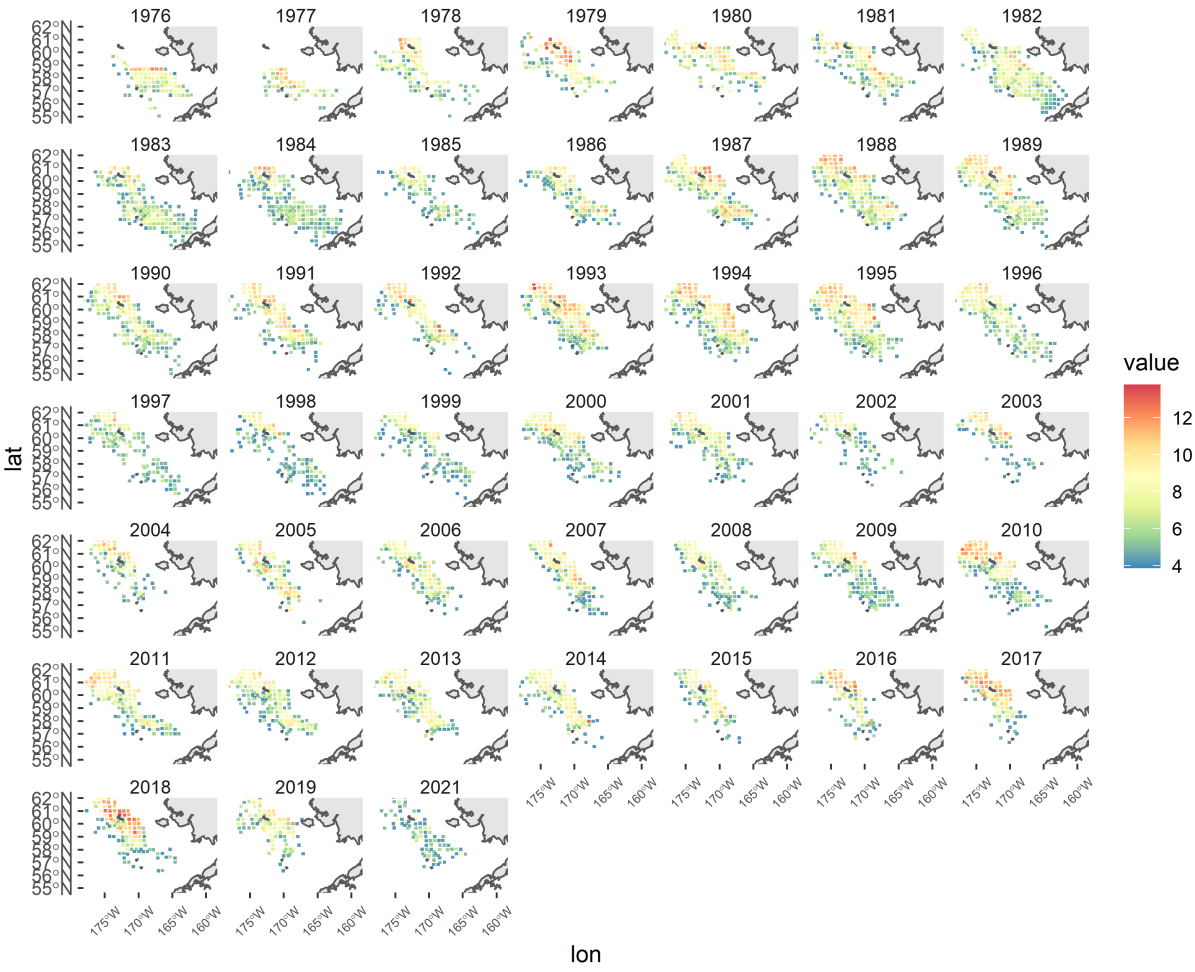

Figure 21: Distribution and intensity of densities (in log numbers) of crab <55 mm carapace width in the NMFS summer survey.

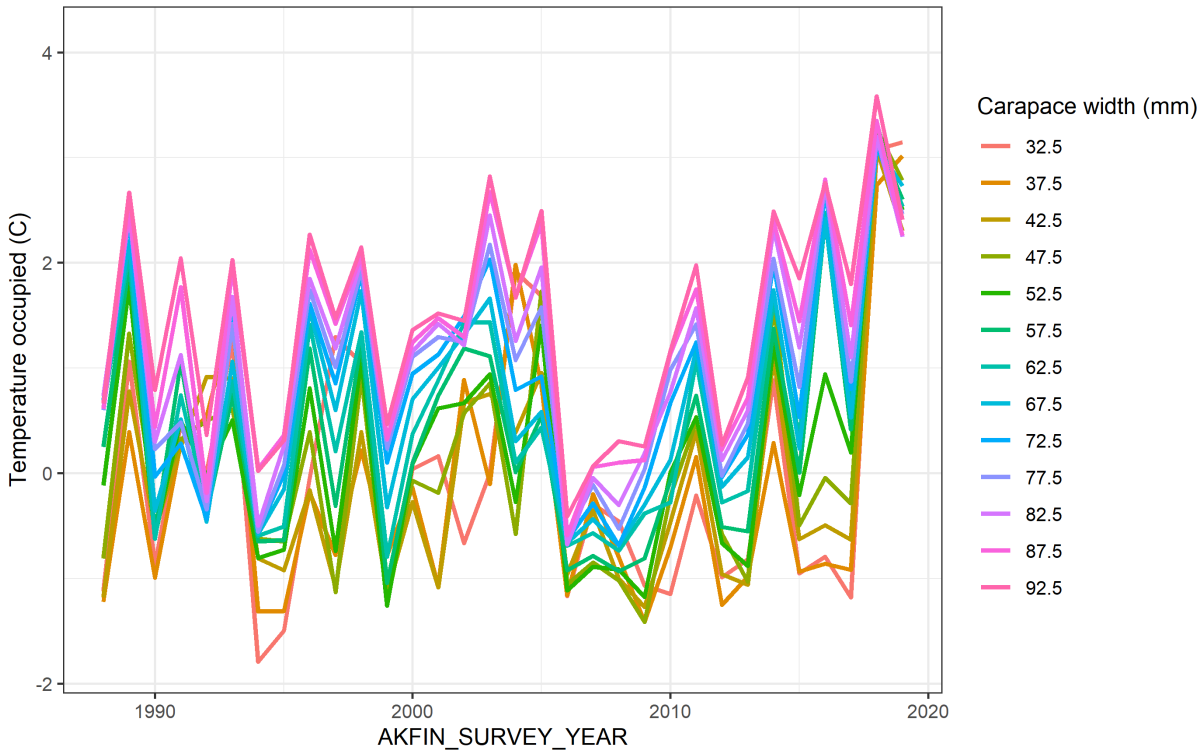

Figure 22: Temperature occupied over time of crab by 5 mm size bin.

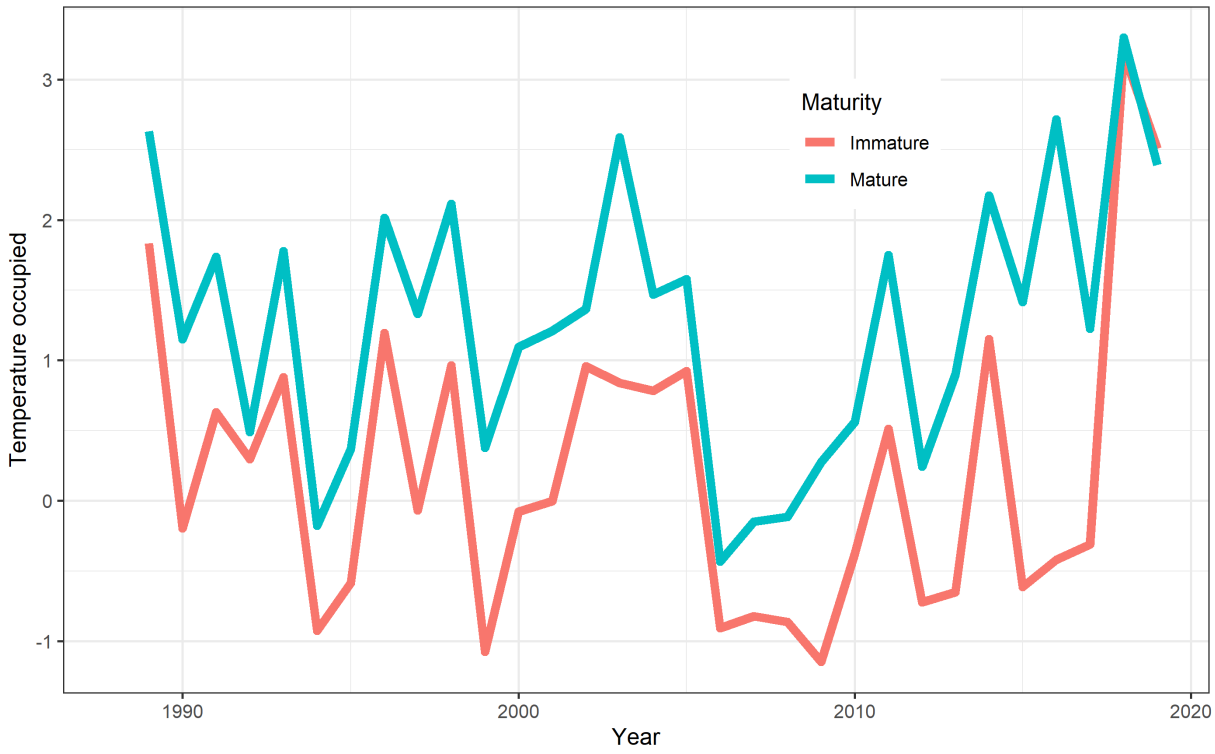

Figure 23: Temperature occupied over time of crab by maturity state defined by the size at which 50% of crab are mature according to chela height.

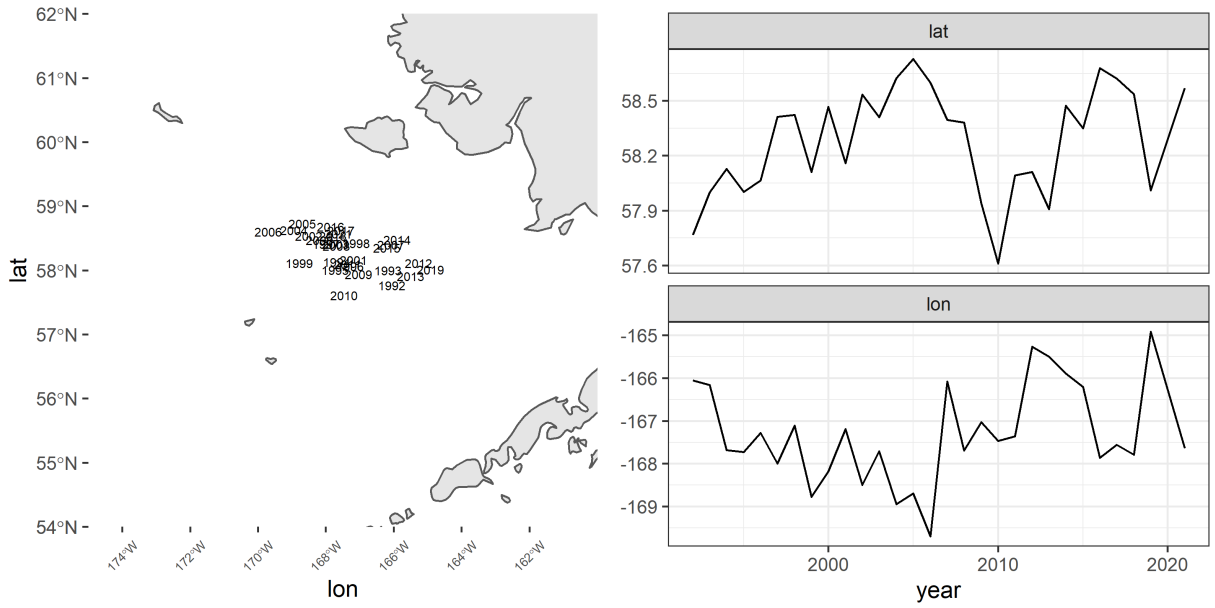

Figure 24: Centroids of abundance for Pacific cod in the Bering Sea over time (left). Right panels show the time series of the centroids broken down by latitudinal and longitudinal components.

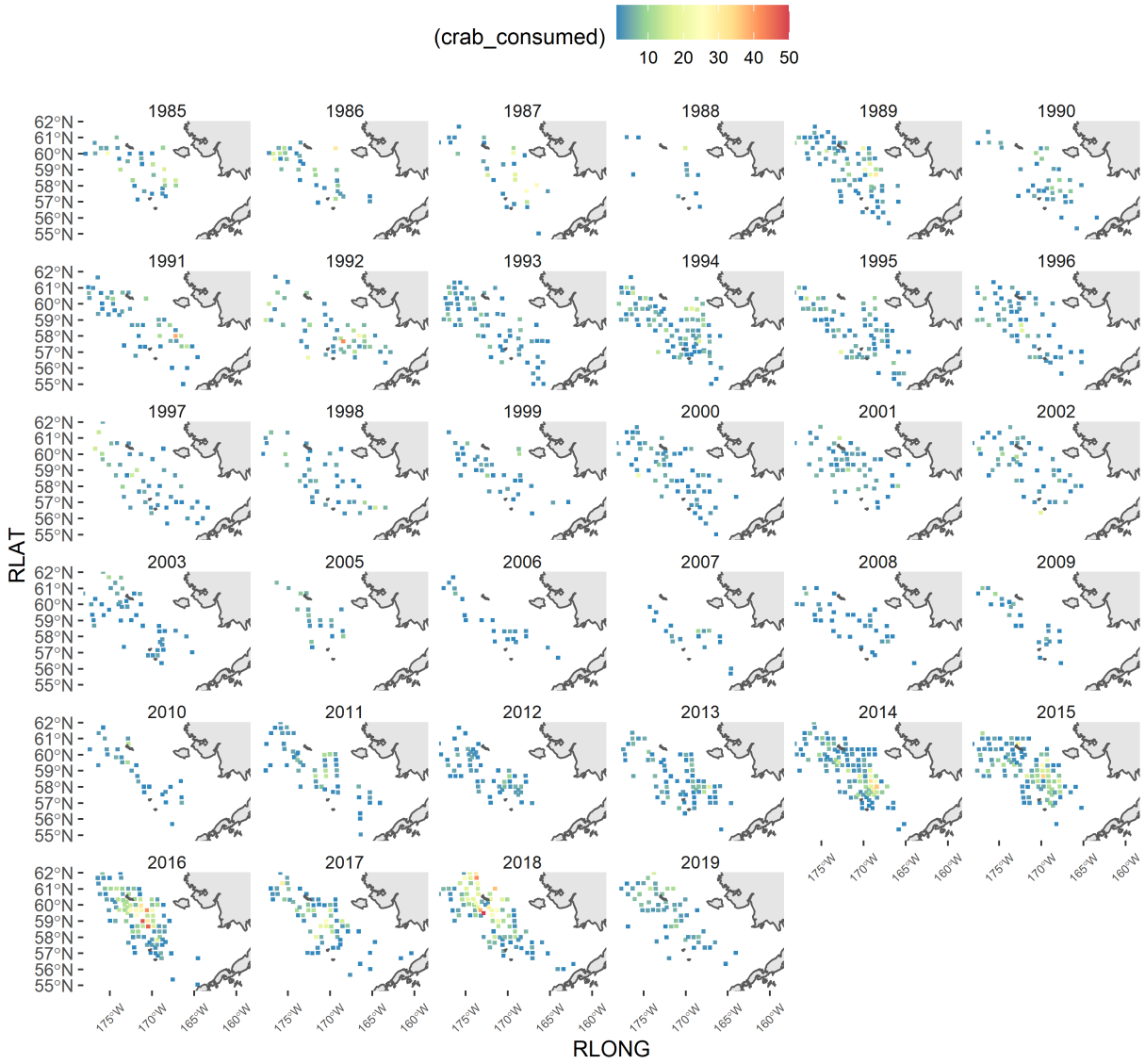

Figure 25: Location and number of crab observed in cod stomachs over time. These are the raw data used to calculate crab consumption by cod and have not been adjusted for sampling effort, but provide background for the spatial distribution of predation over time.

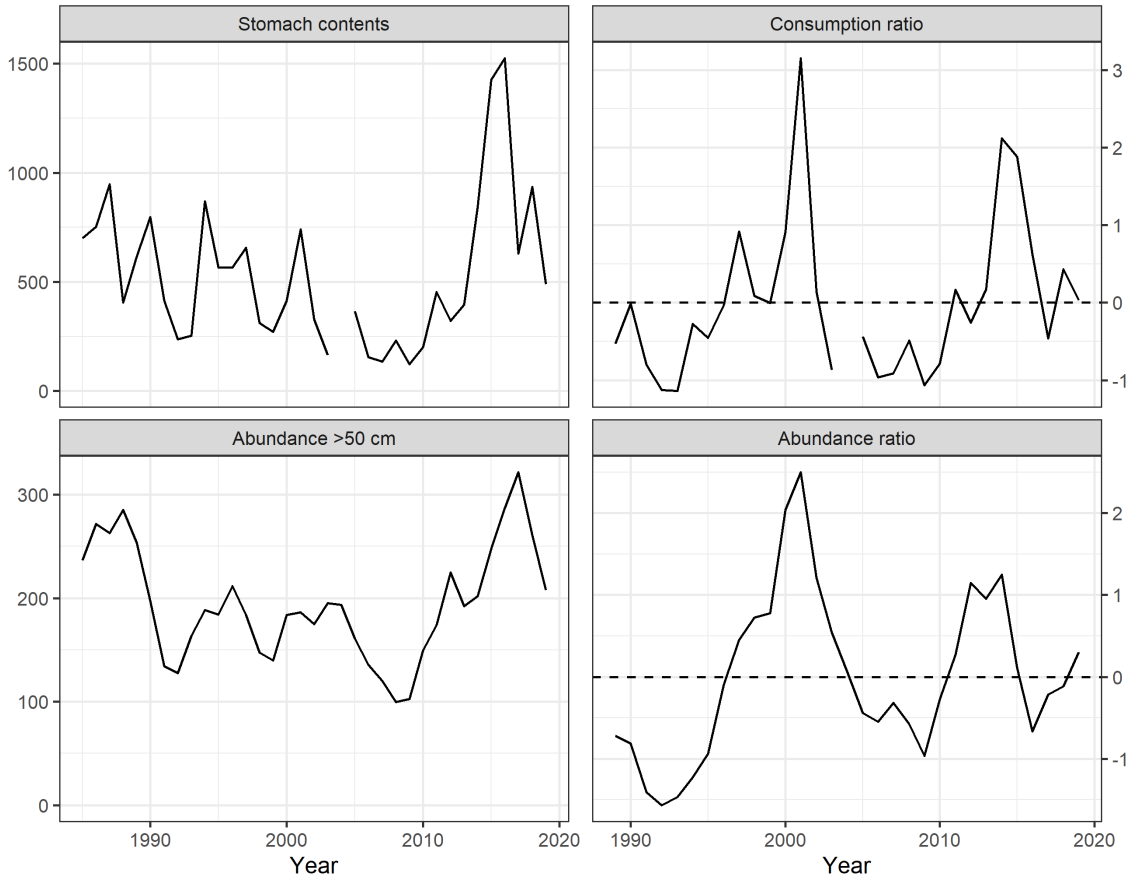

Figure 26: A comparison of indices of cod predation on snow crab. Left column are the calculated consumption of crab by cod (top) and the raw numbers of cod greater than 50 cm. The right column is the left column divided by the estimated number of crab in the eastern Bering Sea.

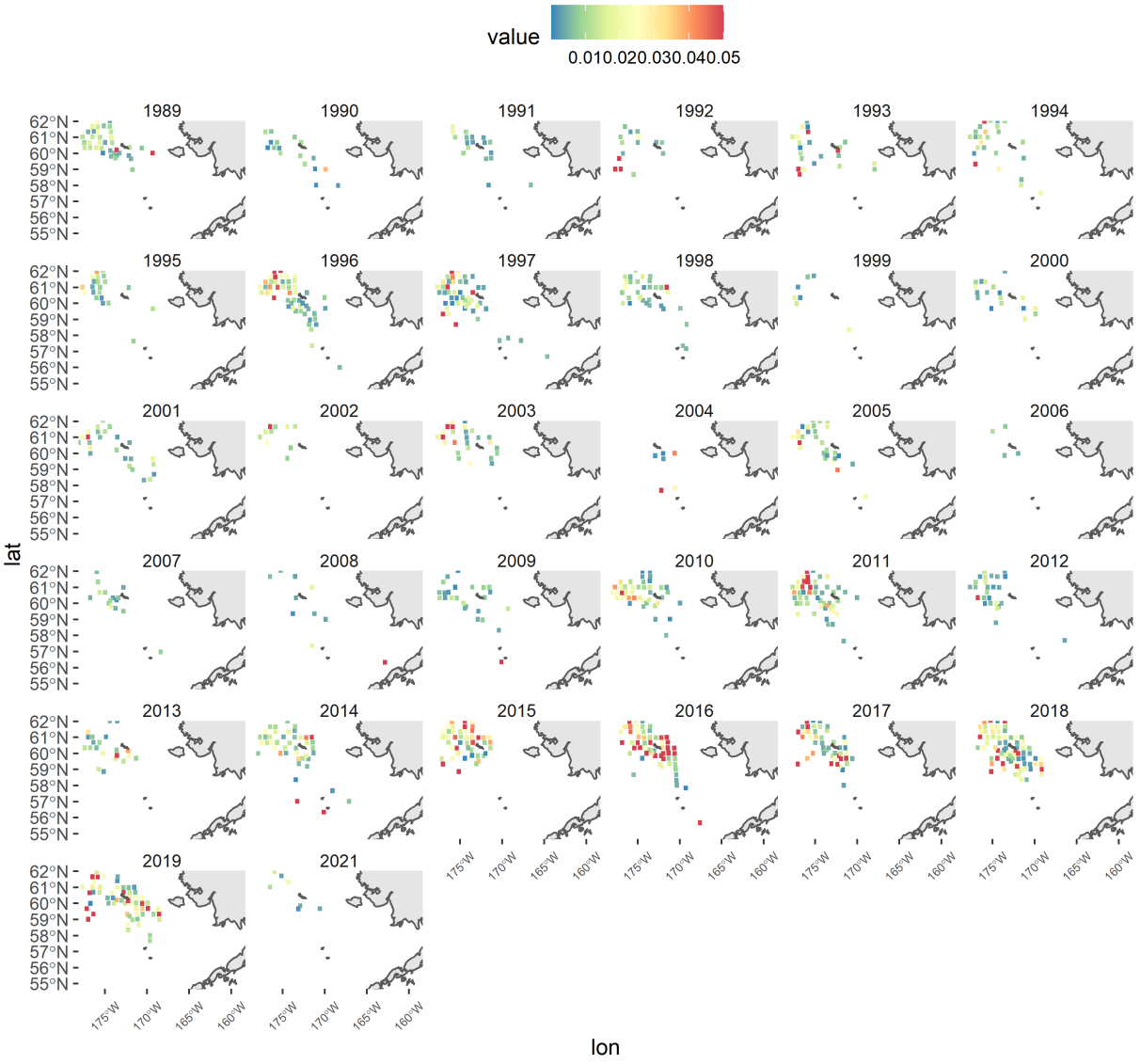

Figure 27: Location and intensity of bitter crab disease over time from visual prevalence observations in the NMFS summer survey.

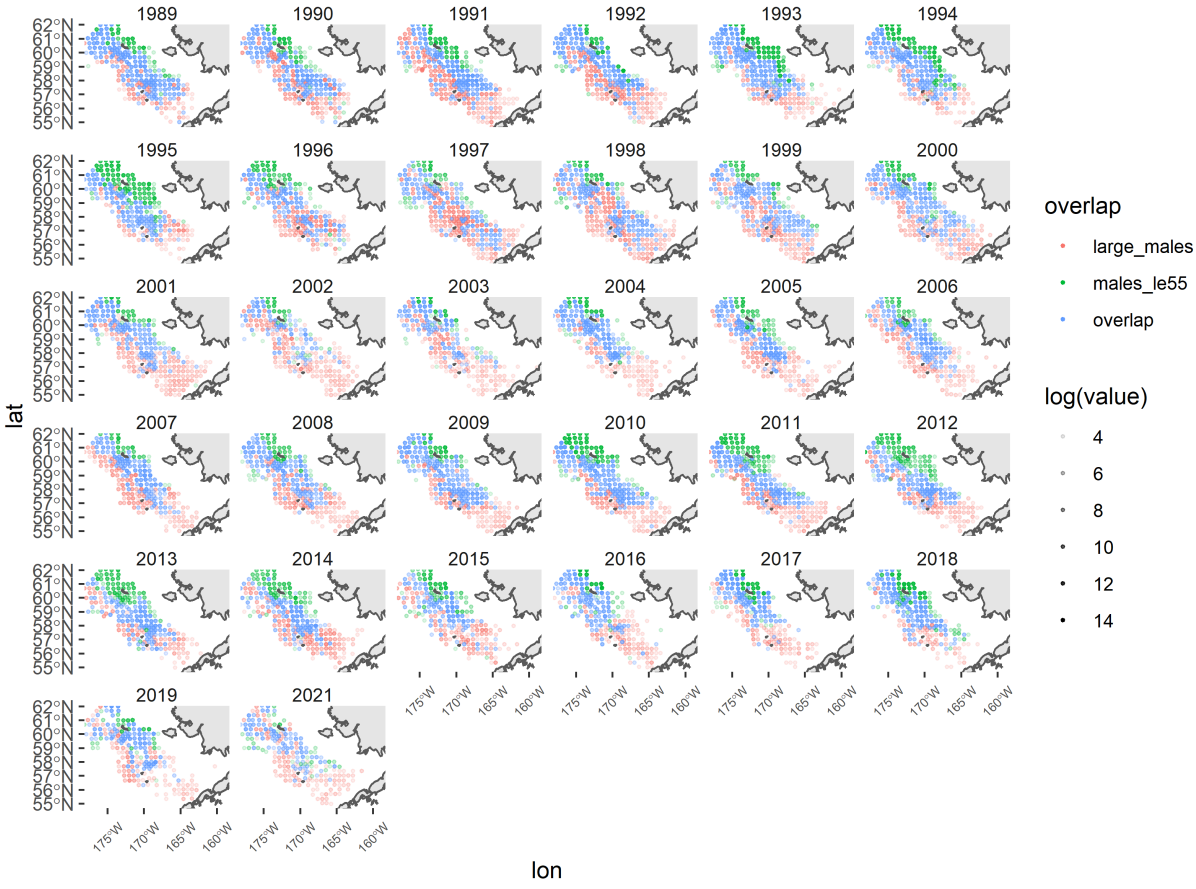

Figure 28: Overlap of large males (>95 mm carapace width) and males smaller than 55 mm carapace width. Opacity of the dot represents the density of crab. Blue represents overlapping distribution. Green and red represent non-overlapping observations of small and large males, respectively.

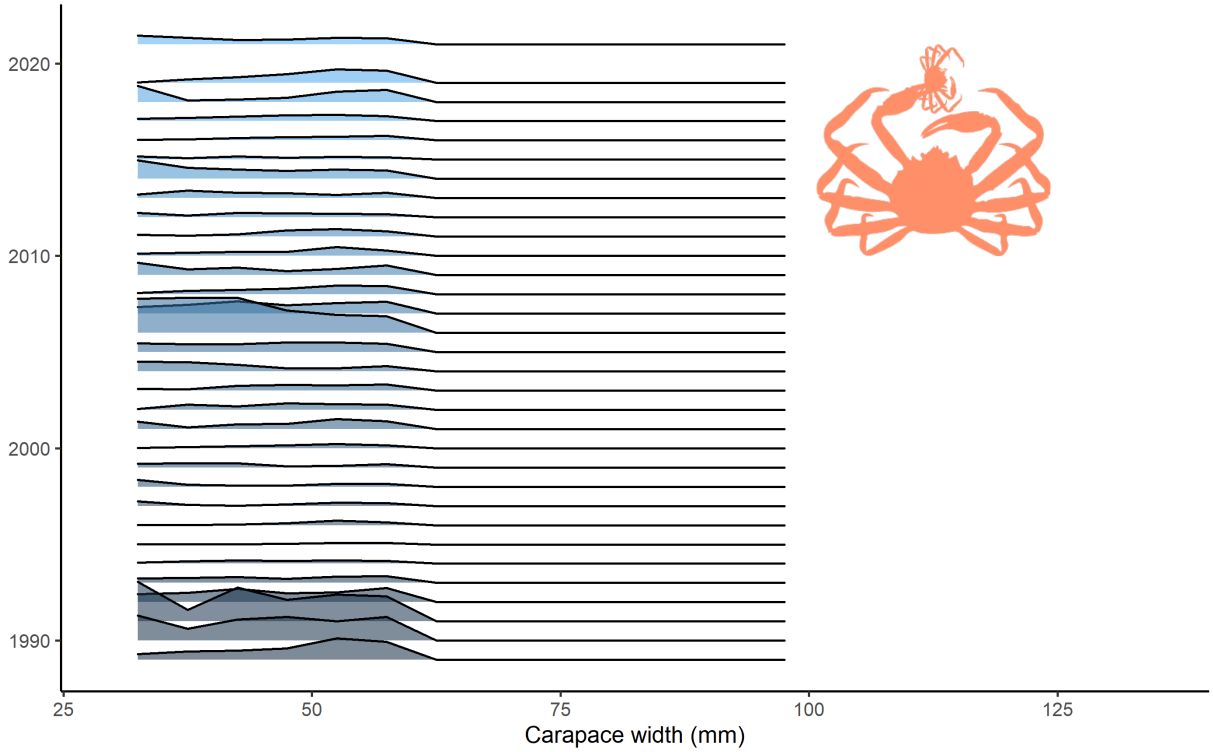

Figure 29: Relative risk at size for cannibalism over time.

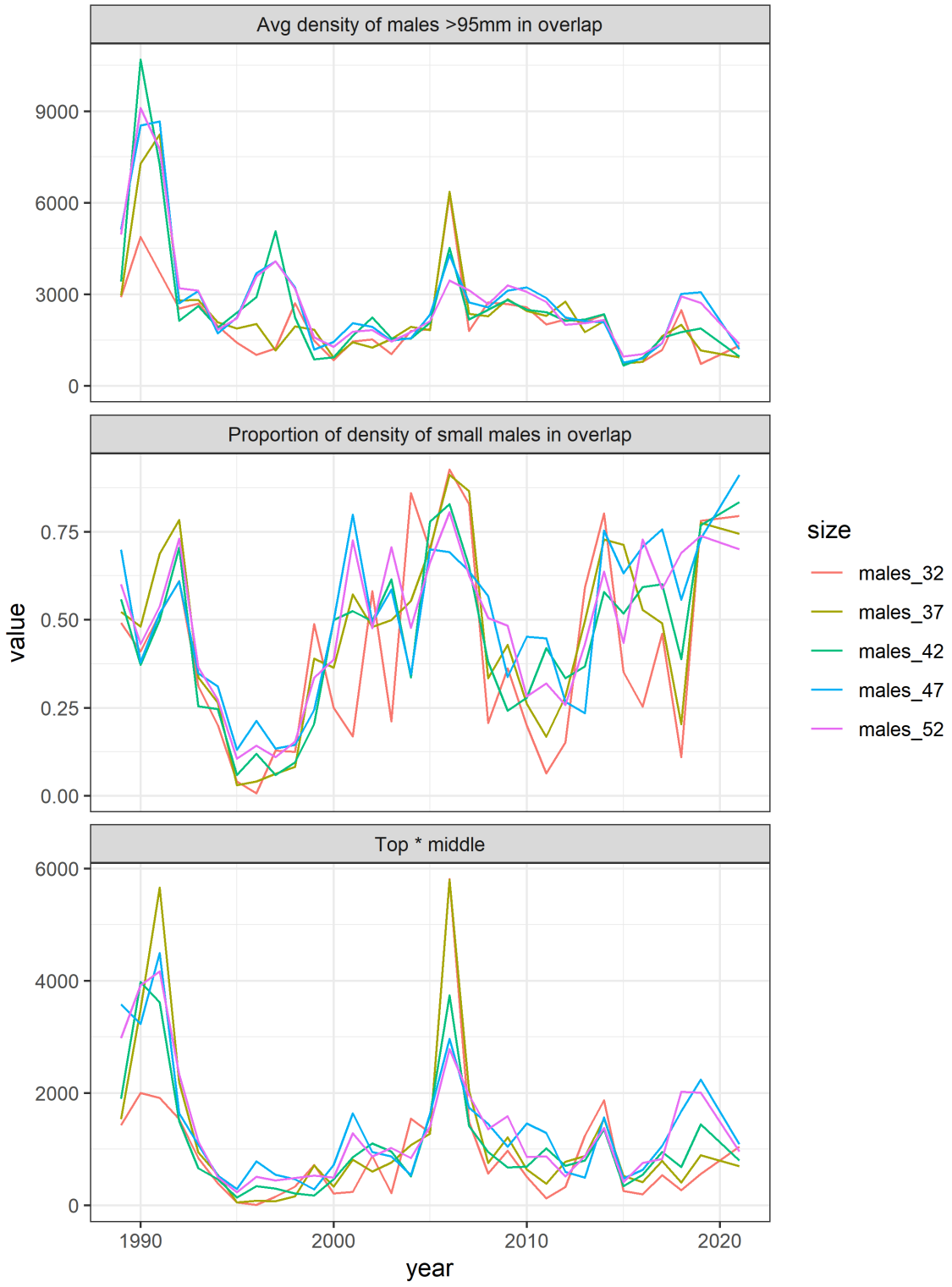

Figure 30: Times series by size of the density of large males in overlapping space (top), the propotion of small males in the overlapping area (middle), and the product of the two (bottom), which is used as an index of cannibalism in the models relating estimated mortality to environmental stressors.

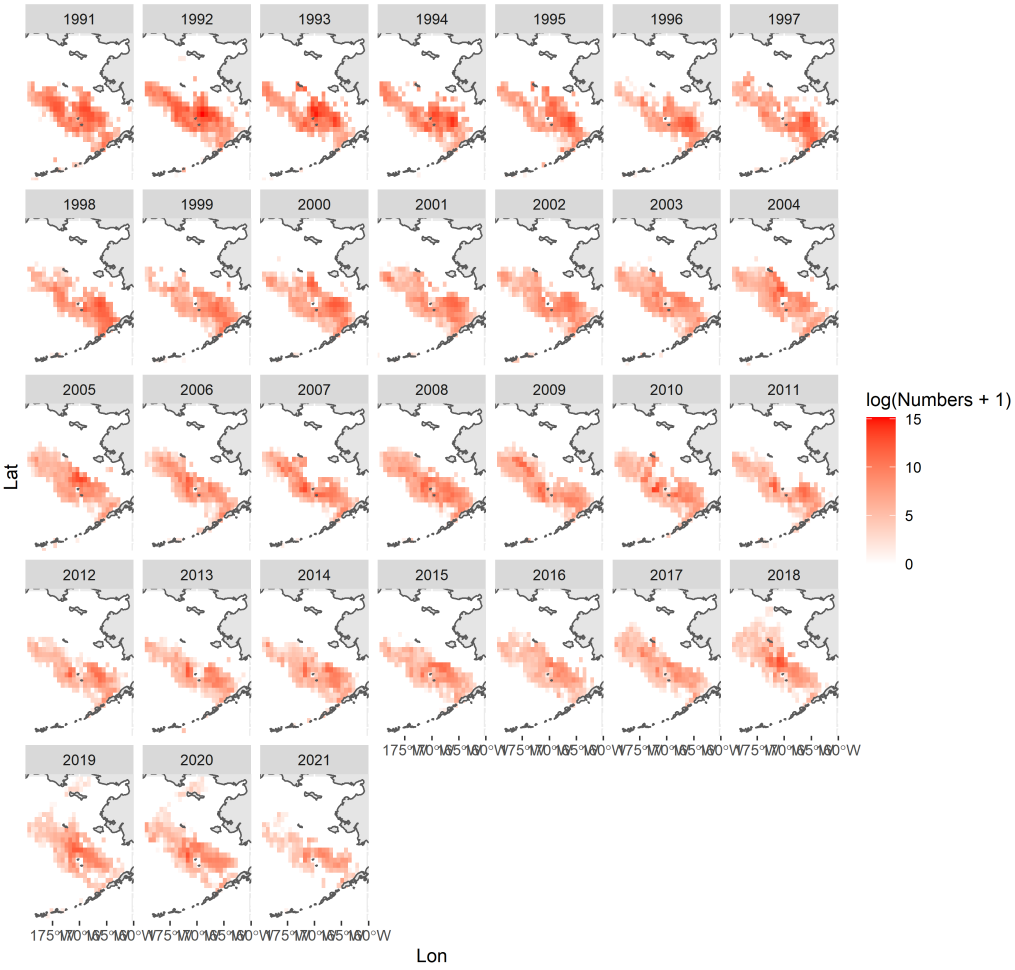

Figure 31: Location and intensity of bycatch of snow crab over time in log space.

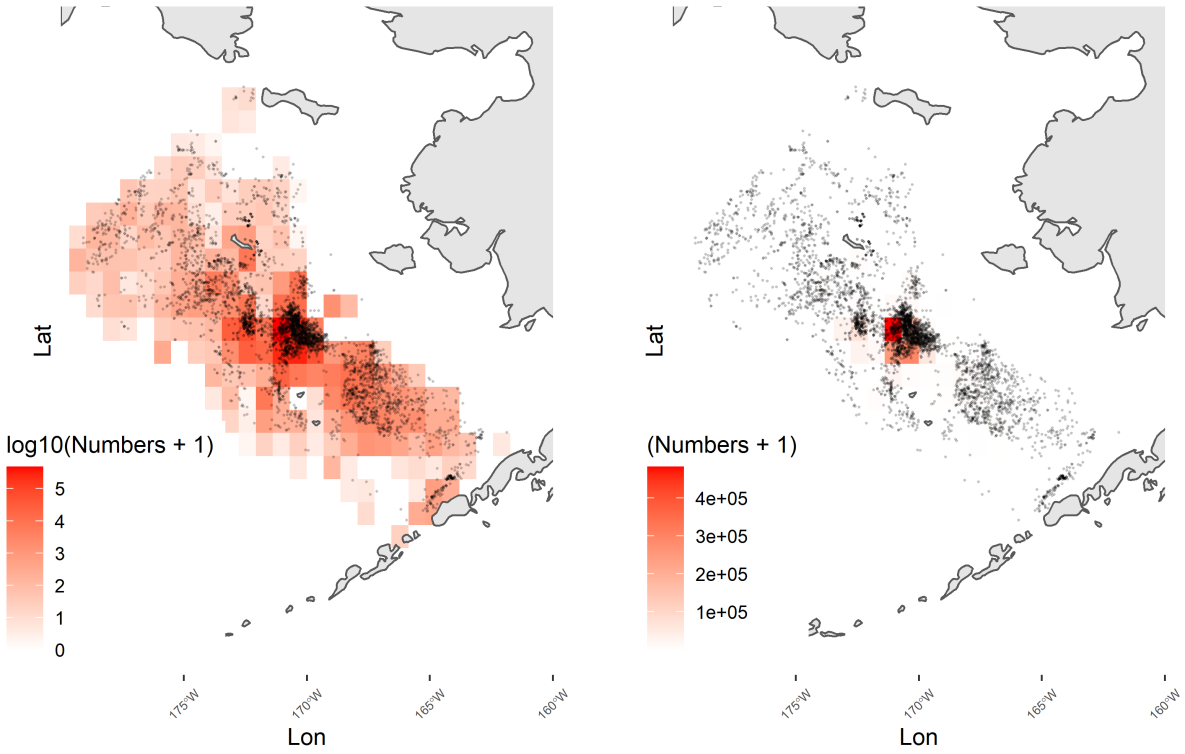

Figure 32: Comparison of location and intensity of bycatch in 2018 for natural and log space.

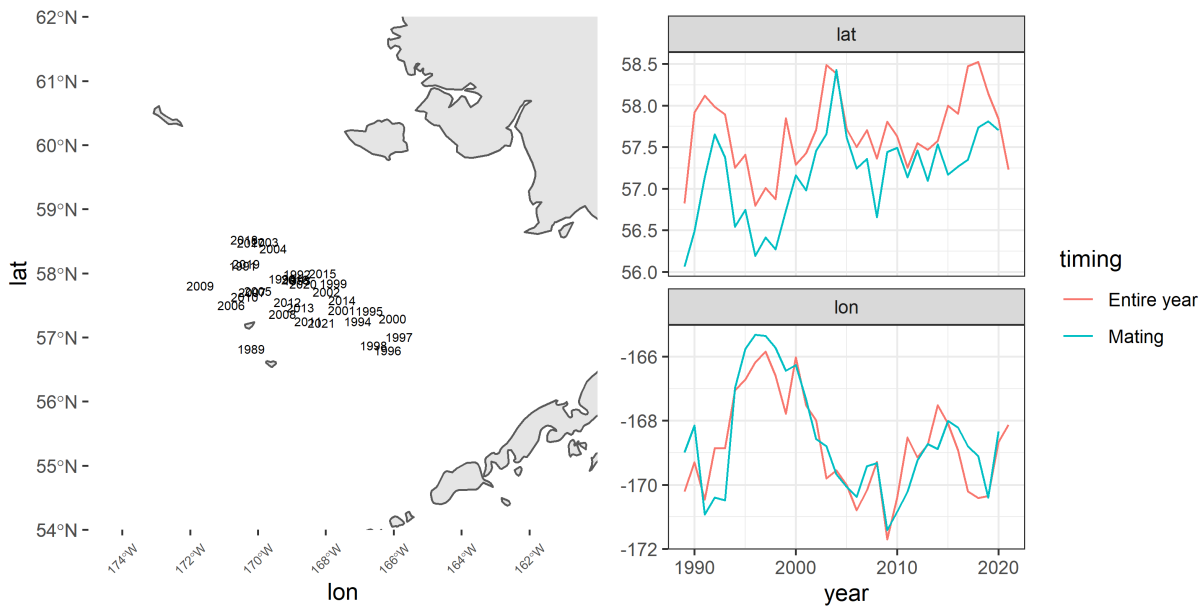

Figure 33: Centroids of bycatch over time calculated over the entire year (left). Centroids broken into time series of latitudinal and longitudinal components calculated over the entire year and during the months December through March which roughly overlap with mating.

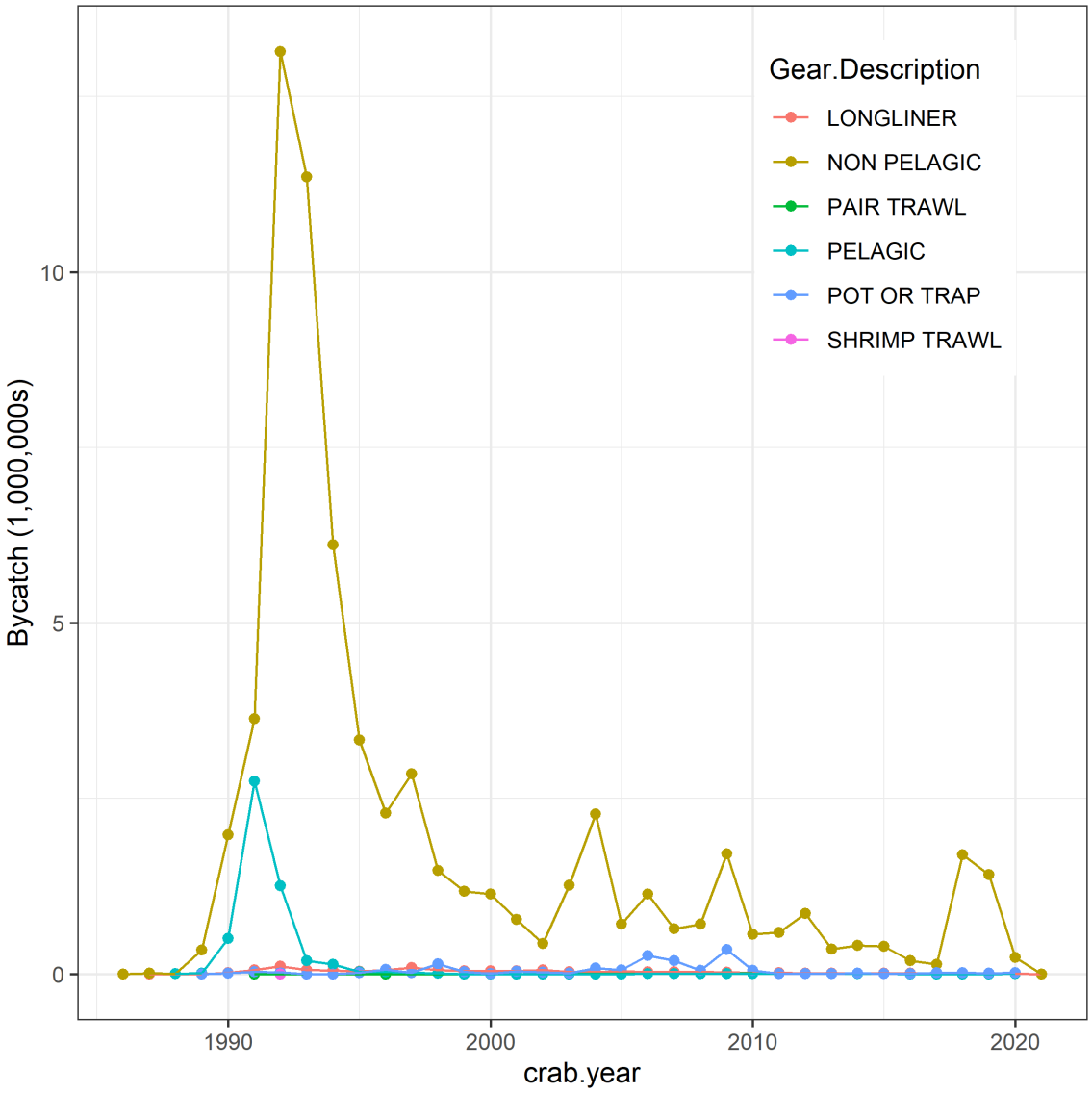

Figure 34: Bycatch in numbers by gear types reported from NMFS observer programs.

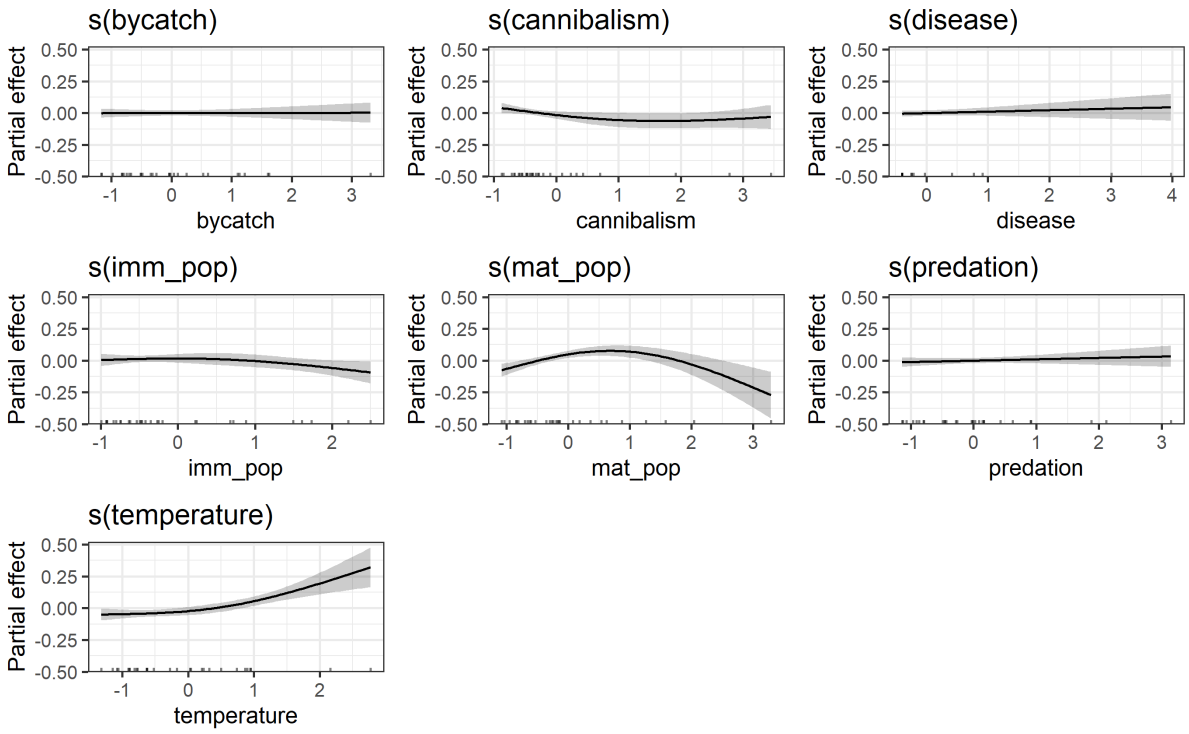

Figure 35: Smooths resulting from the full model estimating the relationship between environmental covariates and immature mortality.

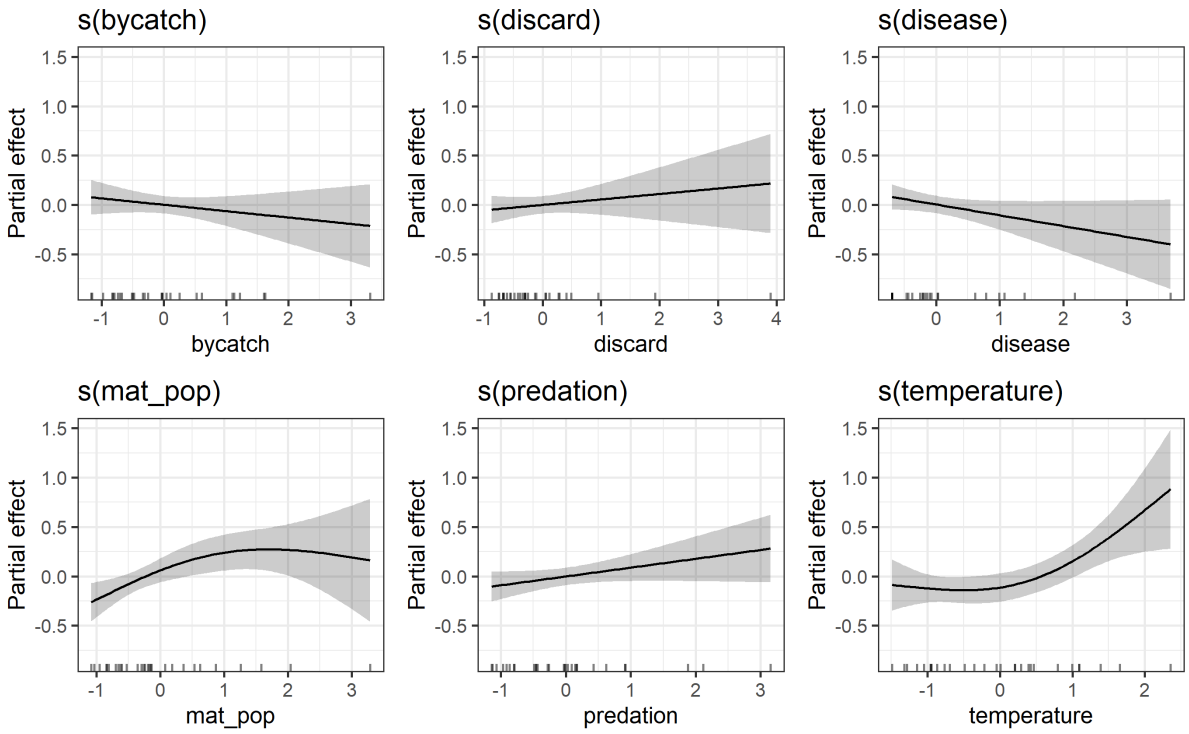

Figure 36: Smooths resulting from the full model estimating the relationship between environmental covariates and mature mortality.

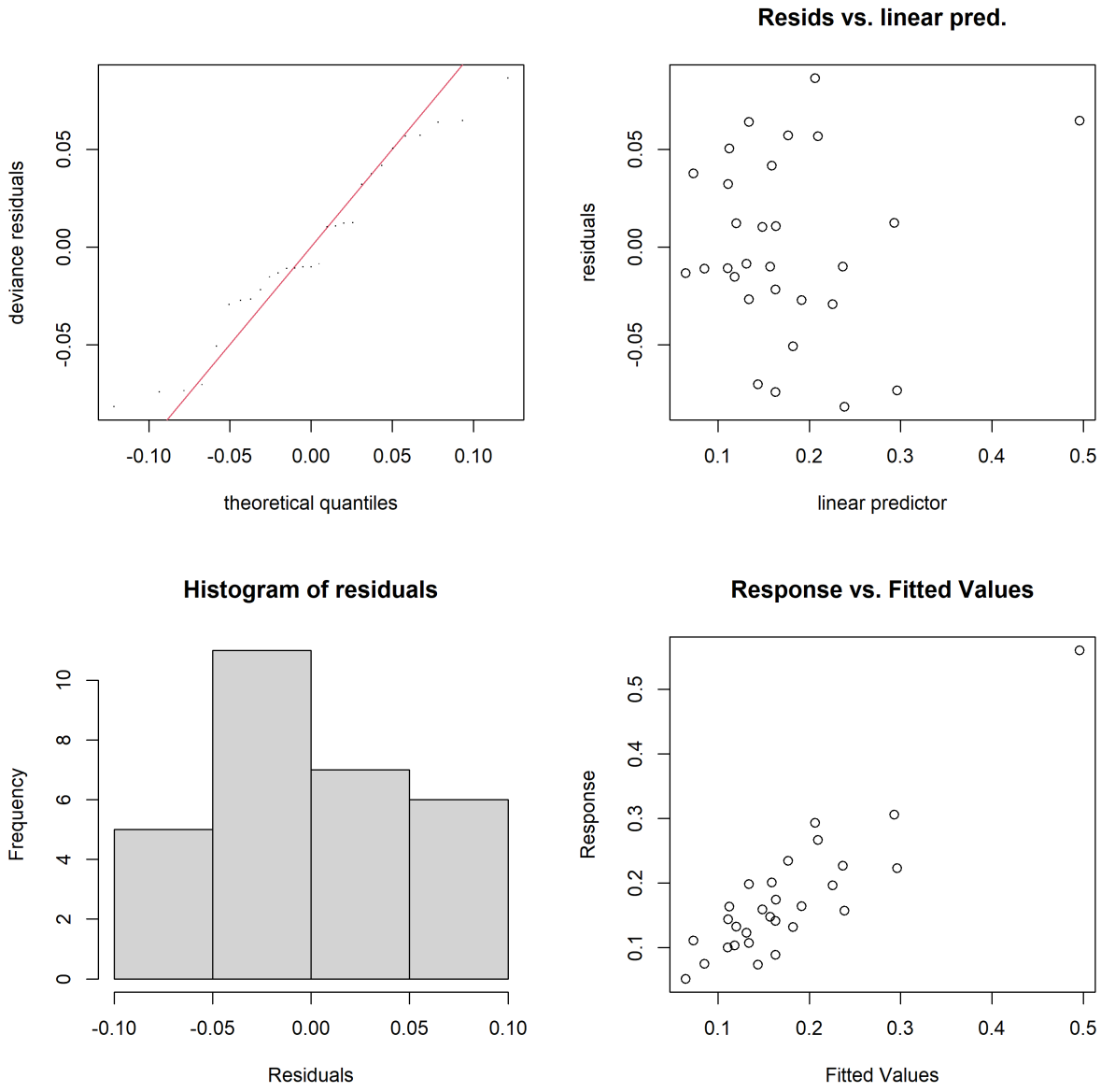

Figure 37: Diagnostic plots for the full model relating immature mortality and environmental stressors.

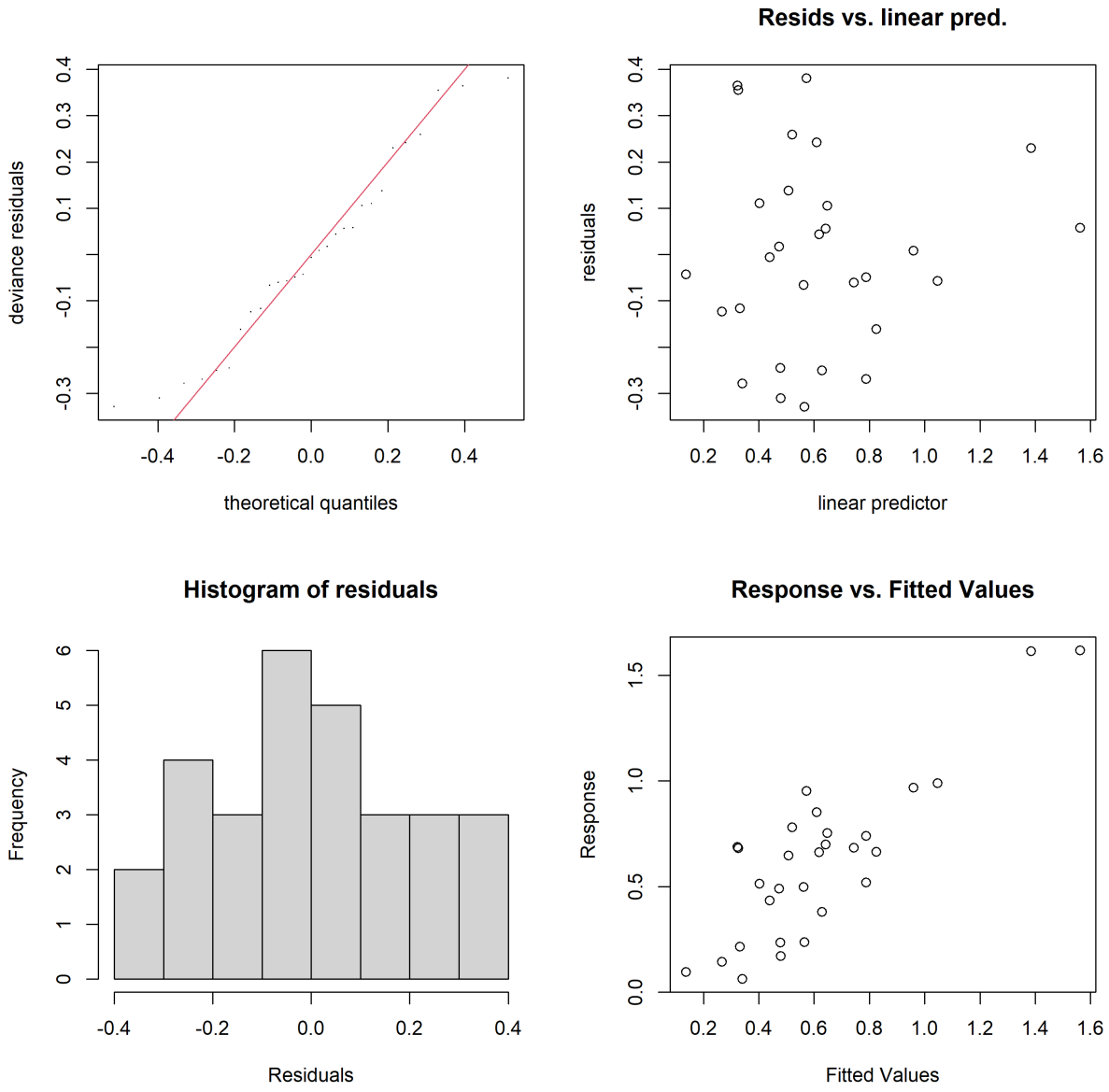

Figure 38: Diagnostic plots for the full model relating mature mortality and environmental stressors

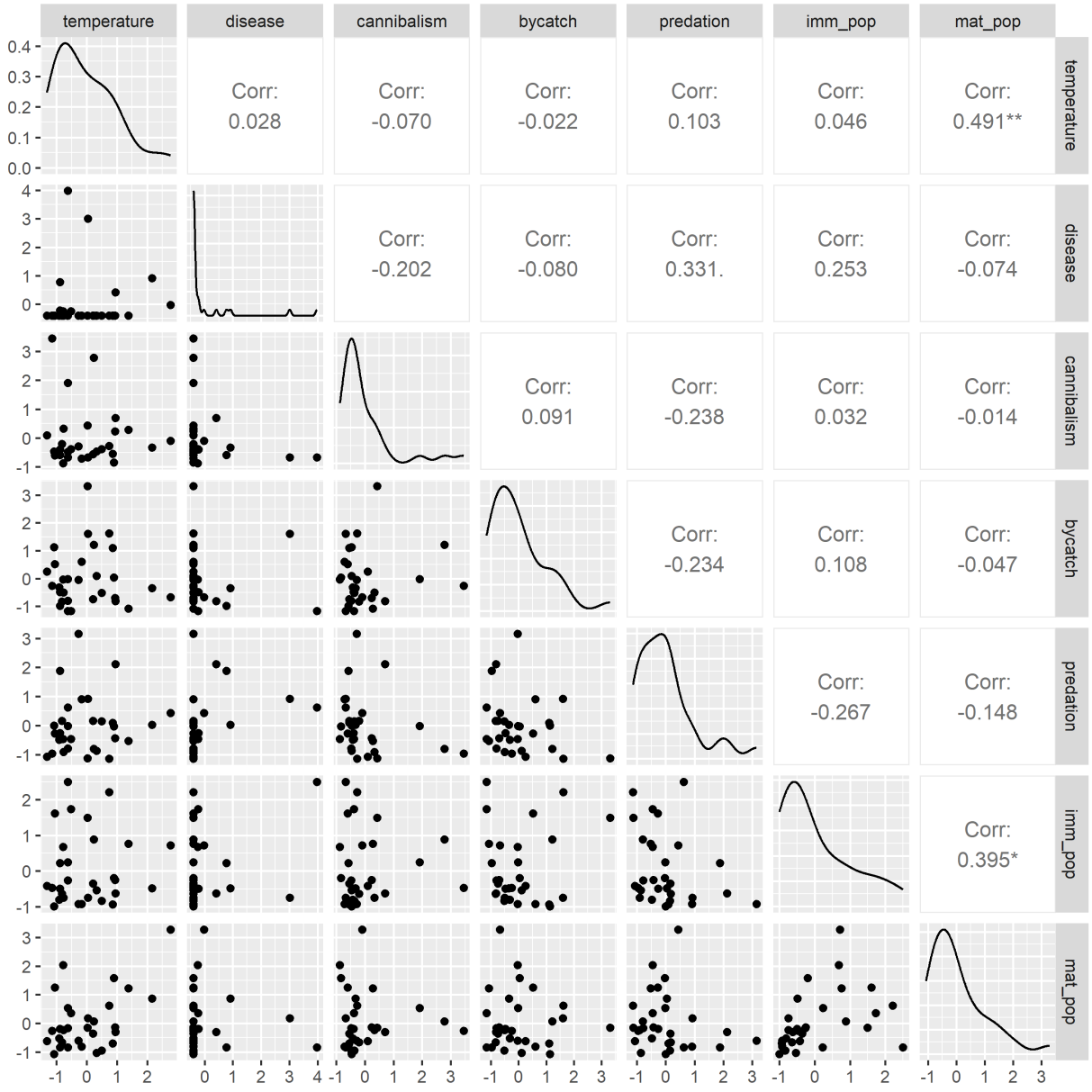

Figure 39: Pairs plots displaying the correlation between covariates for immature crab. Diagonal represents the distribution of a given variable.

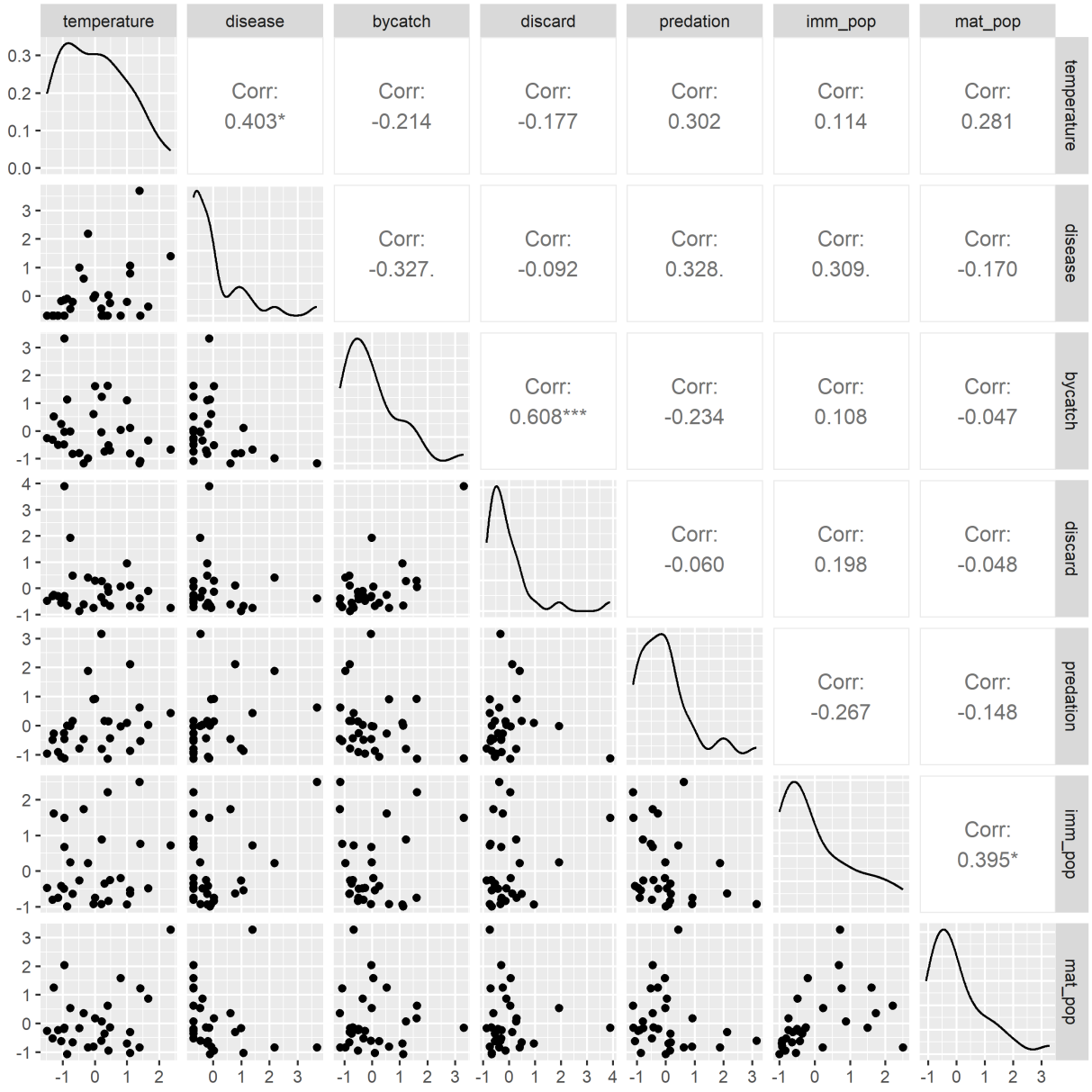

Figure 40: Pairs plots displaying the correlation between covariates for mature crab. Diagonal represents the distribution of a given variable.

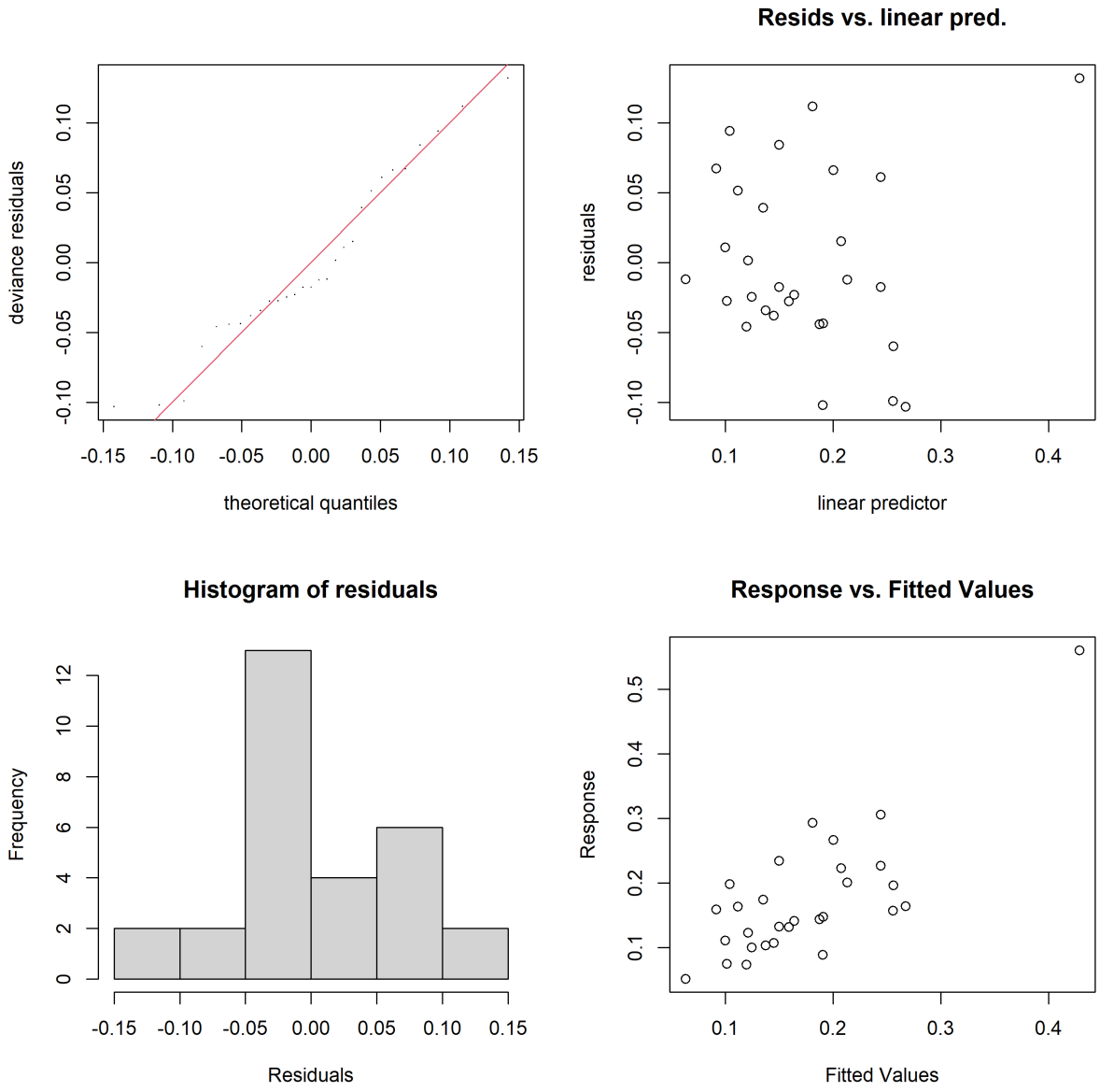

Figure 41: Diagnostic plots for the trimmed model relating immature mortality and environmental stressors.

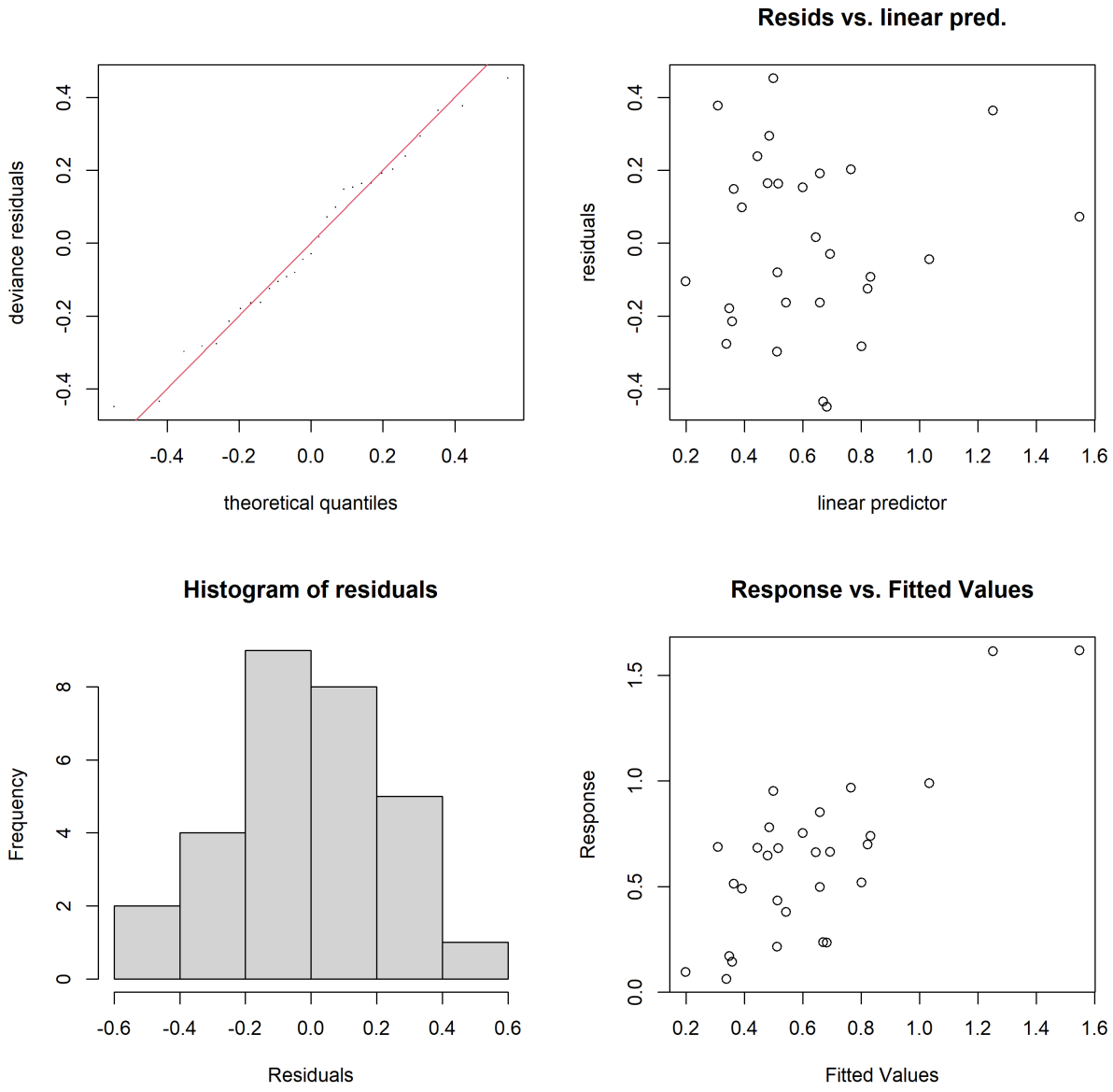

Figure 42: Diagnostic plots for the trimmed model relating mature mortality and environmental stressors.

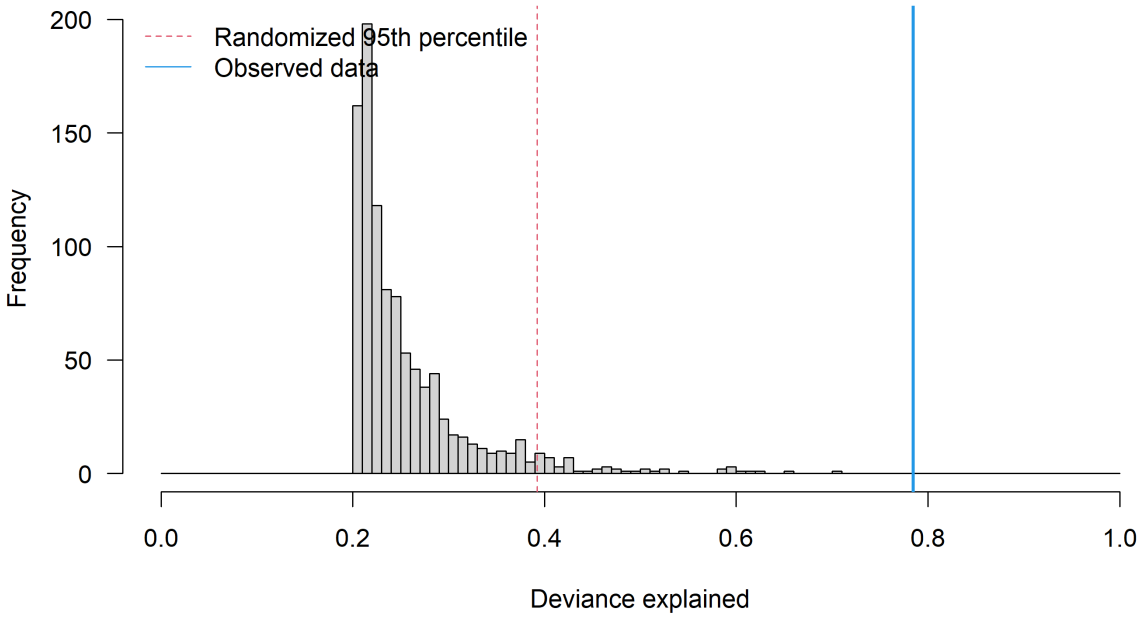

Figure 43: Results of randomization trials for the trimmed model relating estimated immature mortality to environmental stressors. Grey bars represent the number of trials in which the randomized model explained the deviance on the x-axis. Dashed vertical red line represents the 95th quantile of the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

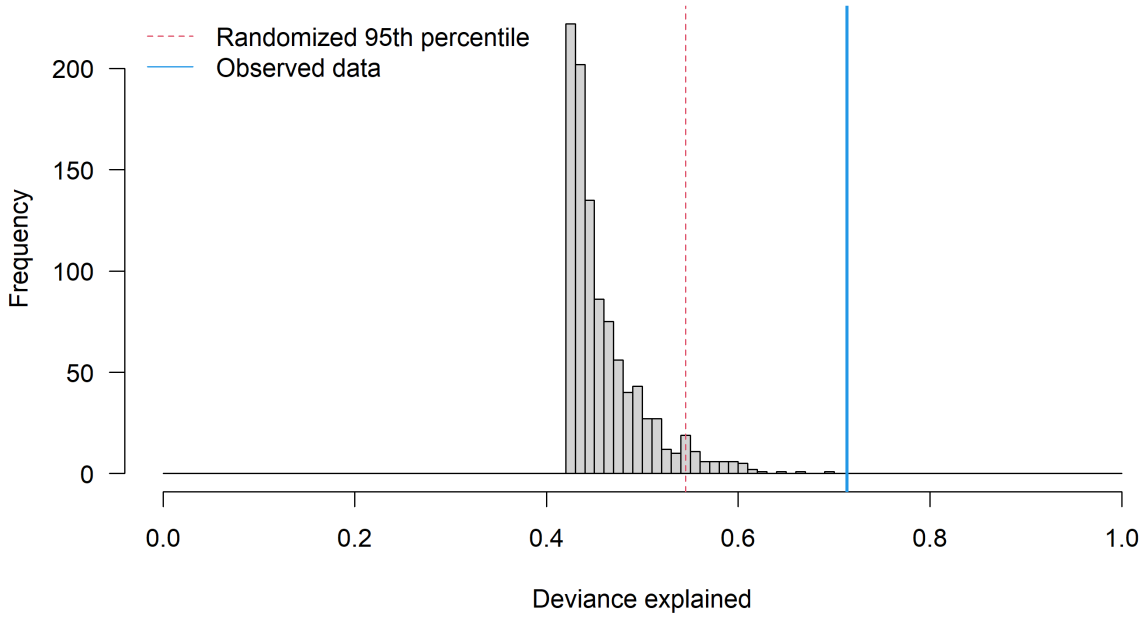

Figure 44: Results of randomization trials for the trimmed models relating estimated immature mortality to environmental stressors. Grey bars represent the number of trials in which the randomized model explained the deviance on the x-axis. Dashed vertical red line represents the 95th quantile of the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

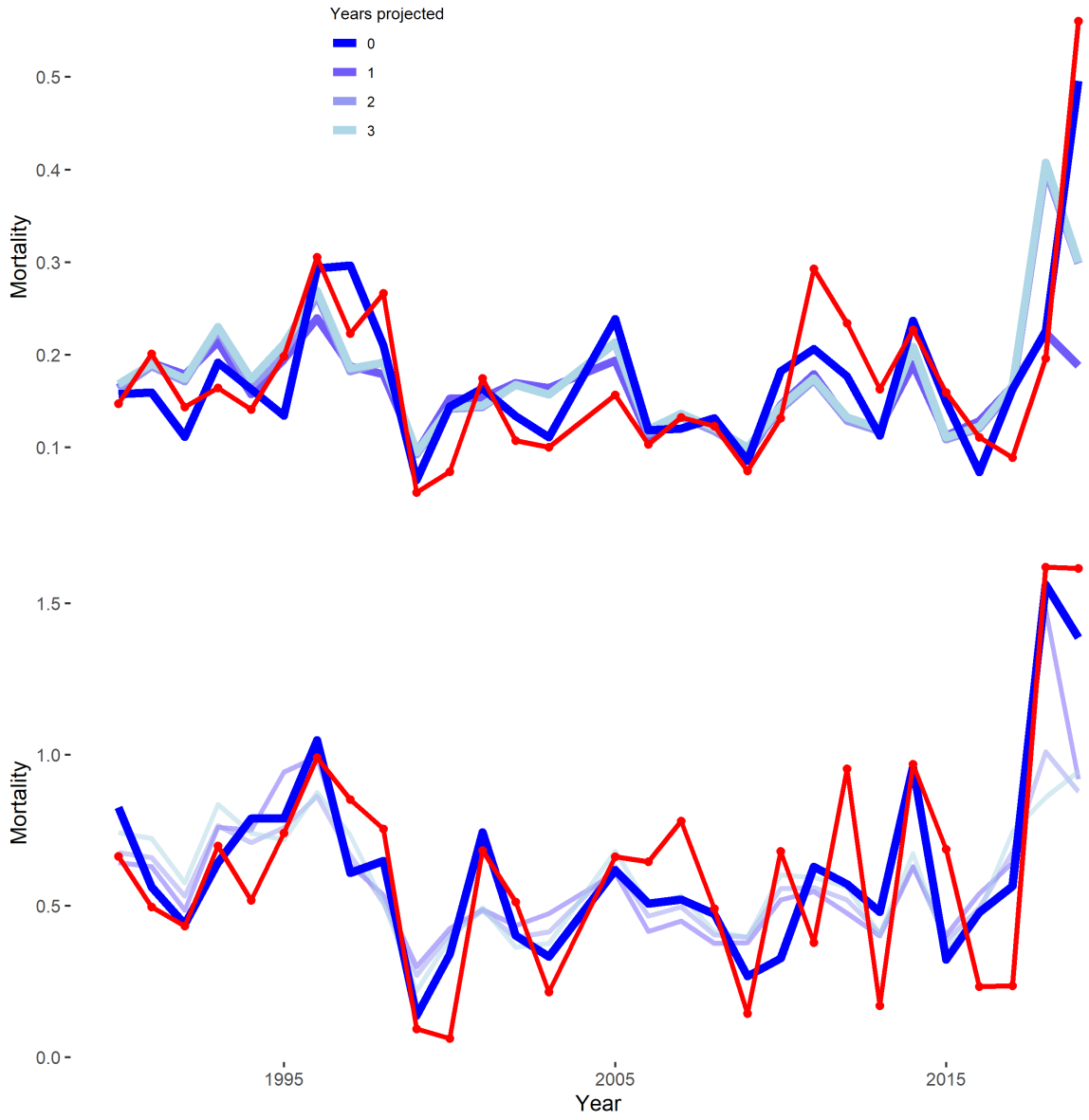

Figure 45: Predictive skill of the GAMs for immature and mature mortality. Reproduced from figure 2 in the main text to provide better detail.

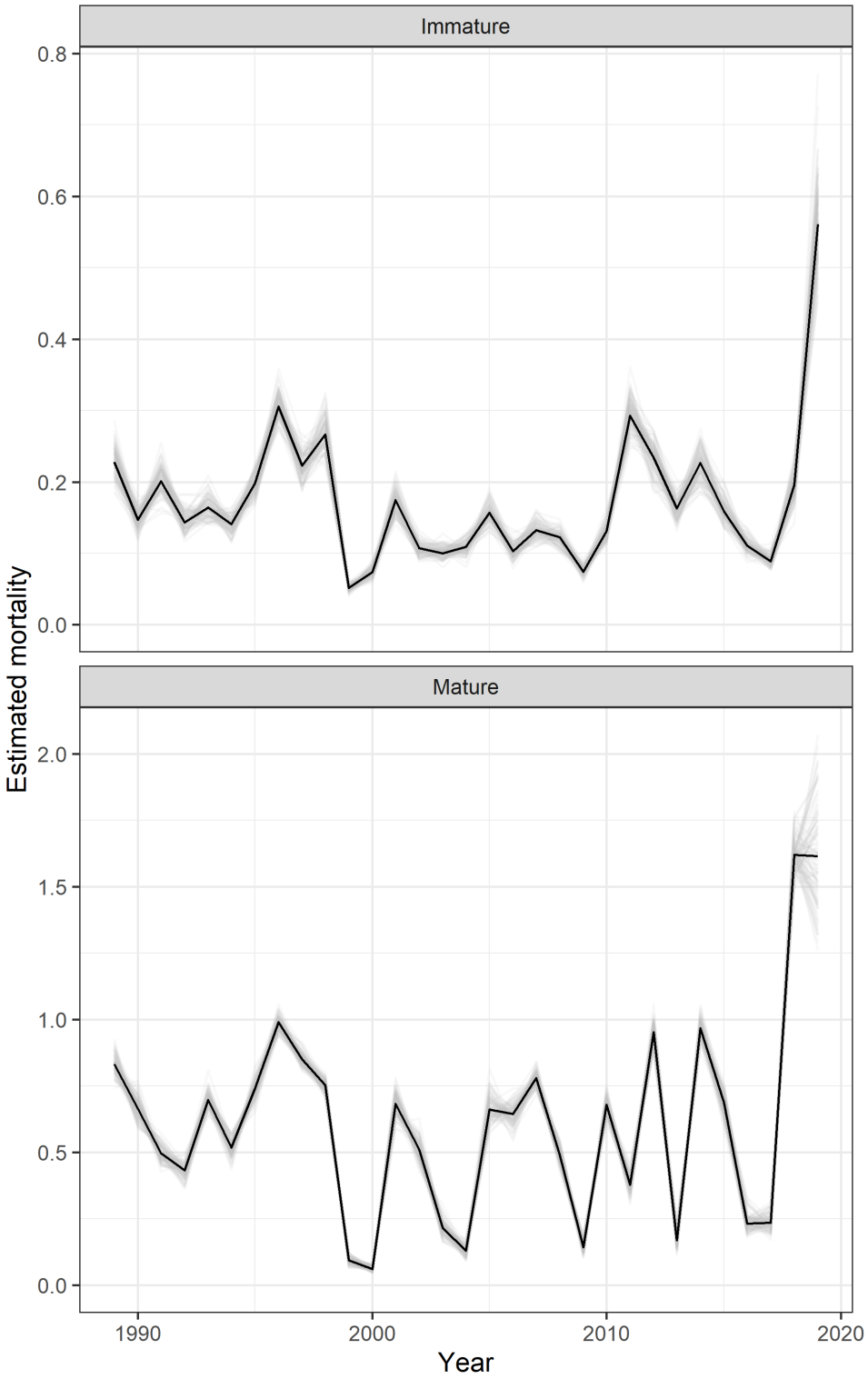

Figure 46: Simulated time series of estimated immature and mature mortality that incorporate the uncertainty associated with the fitting process of the population dynamics model. Each grey line represents one iteration of multiplying the maximum likelihood estimates of the mortality deviations by the covariance matrix. The black line represents the MLE.

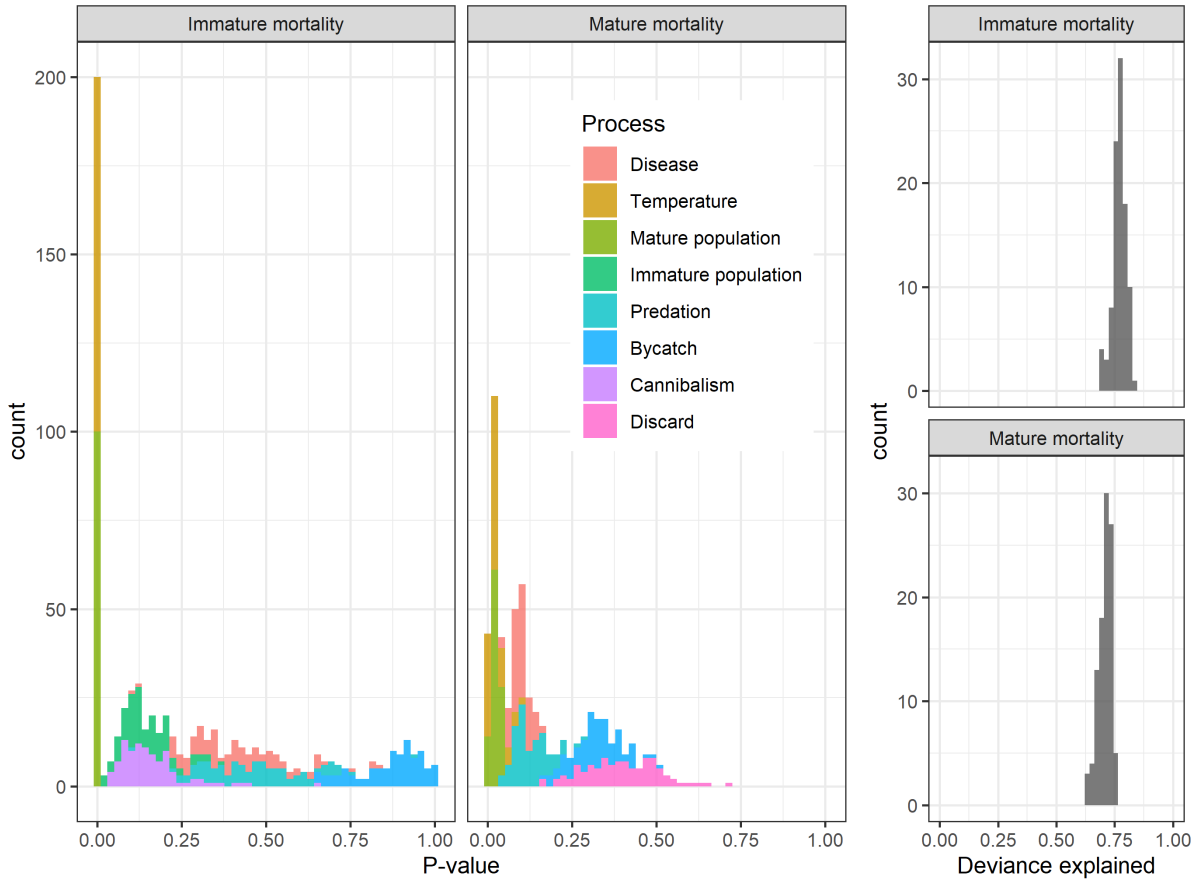

Figure 47: P-values associated with iterations of fitting the GAMs to simulated time series of estimated immature and mature mortality using the covariance matrices estimated in the fitting of the population dynamics model.

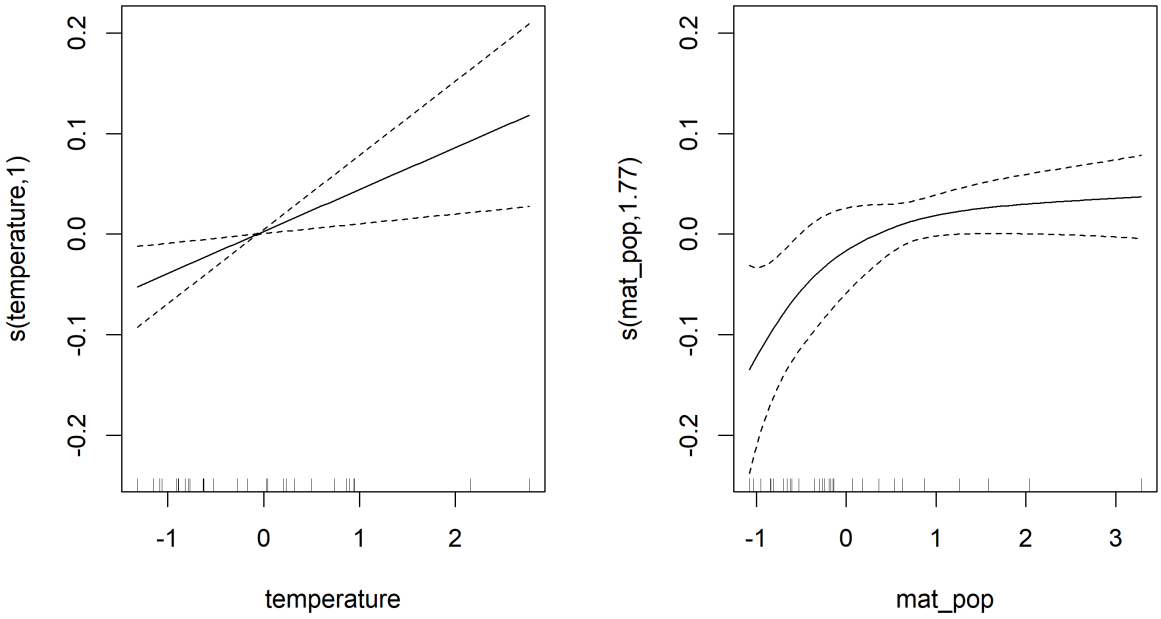

Figure 48: Estimated smooths between immature mortality and temperature occupied and mature population from shape constrained additive models.

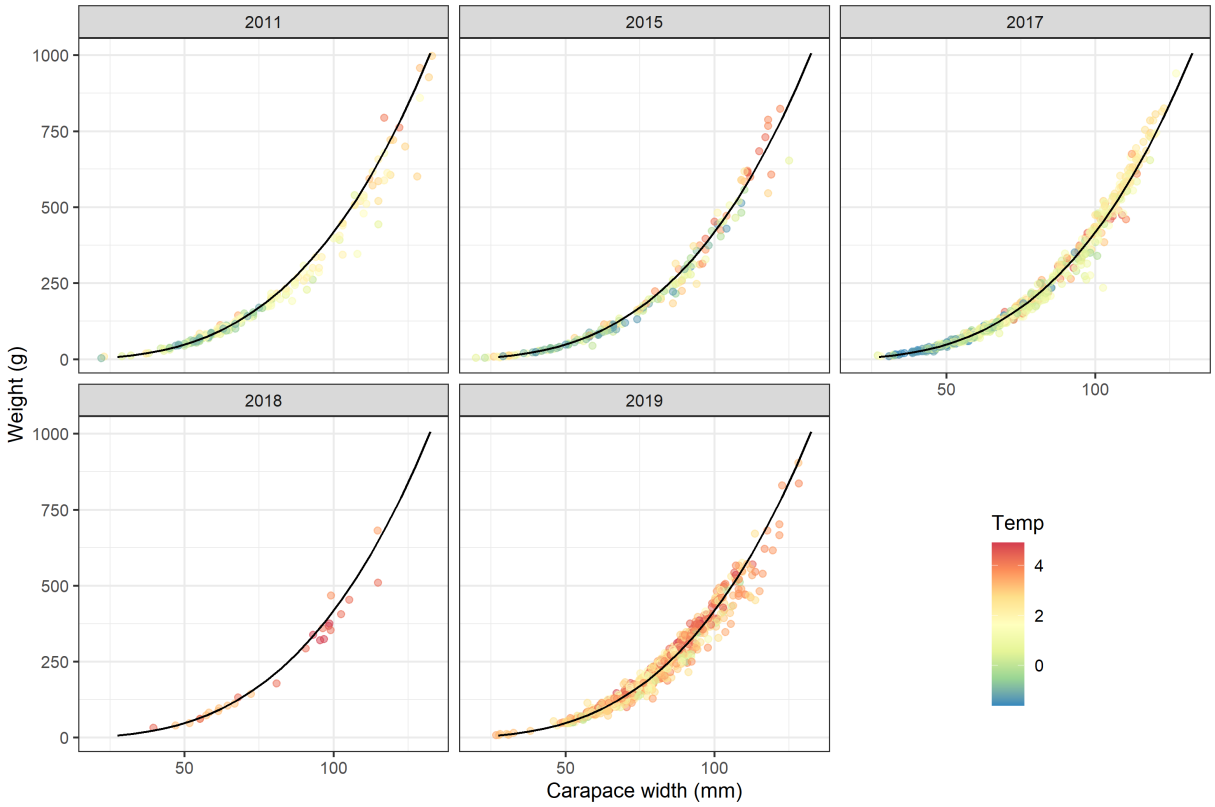

Figure 49: Observed weight at size over time colored by temperature (Celcius) at which the crab was collected.

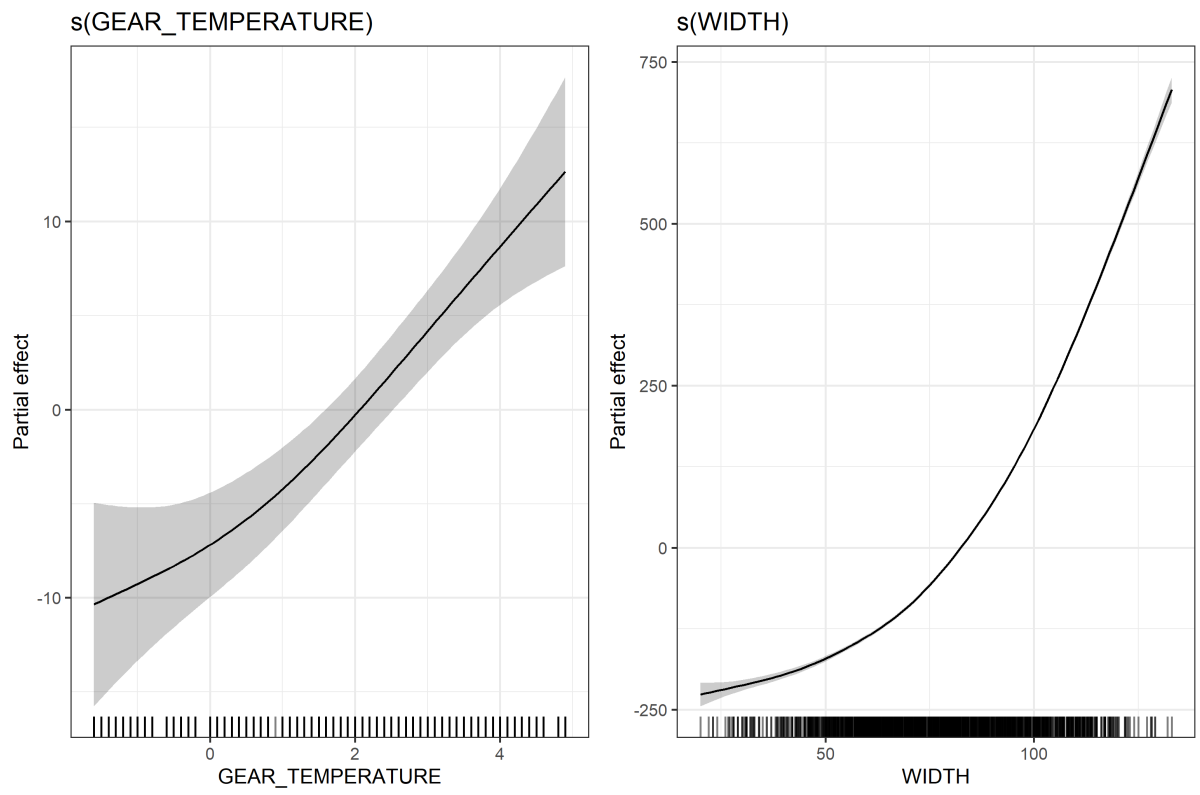

Figure 50: GAM estimated relationships between temperature and carapace width on observed weights of crab.

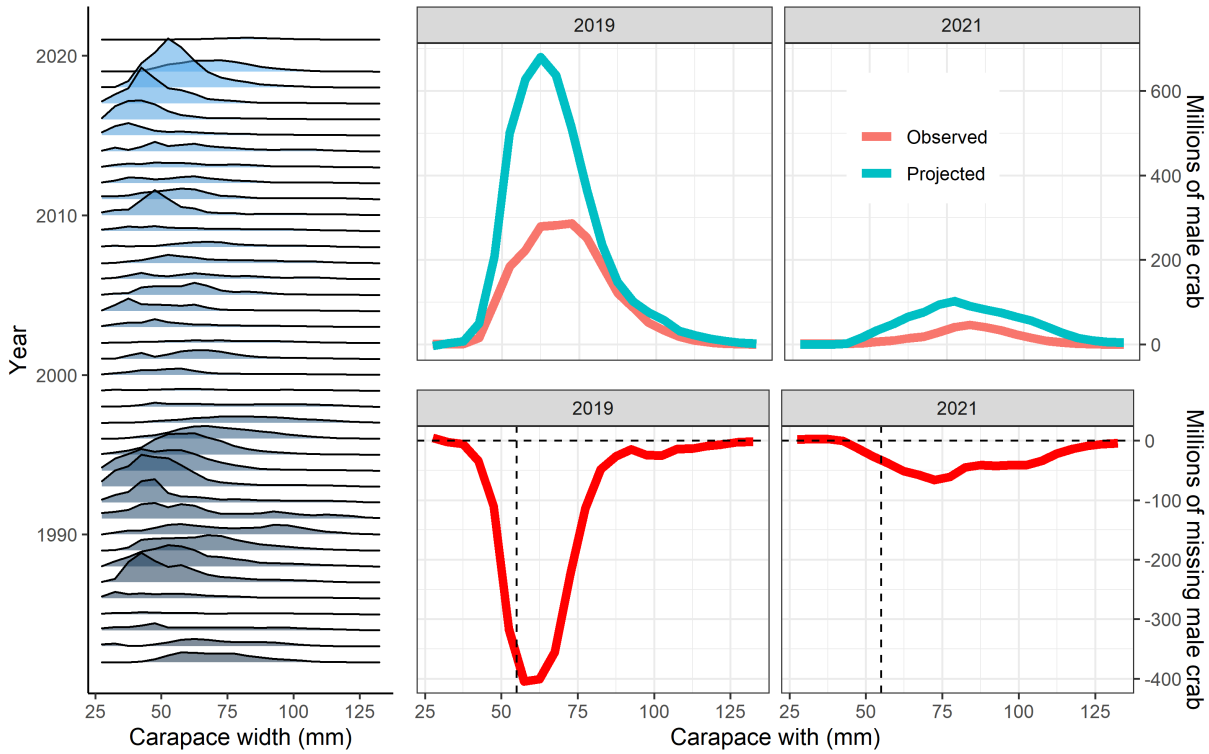

Figure 51: Numbers at size over time of snow crab (left). Observed numbers of crab (red line) in 2019 and 2021 vs. projected numbers of crab from 2018 and 2019 given a mortality equal to 0.27 (the assumed value in the assessment; top left). Numbers of missing crab at size (red line) with the size of crab beneath which cod predate upon (dashed vertical black line).