Estimating time-variation in confounded processes in population dynamics modeling: a case study for snow crab in the eastern Bering Sea

Cody Szuwalski

Population dynamics models used to provide management advice for harvested natural resources often estimate population processes that can be mutually confounded. These processes (e.g., natural mortality (M) and catchability (q)) are often assumed to be time-invariant, but this assumption can be violated in real populations. Not allowing for time-variation in these processes can result in retrospective patterns in estimated biomass, which can result in errors in management advice. Allowing for time-variation in these processes can also result in errors in management advice if the estimated time-variation does not reflect the true underlying variation. Here I present a population dynamics model for snow crab in the eastern Bering Sea that tests different combinations of implementation of time-variation in M and q. The North Pacific Fisheries Management Council did not adopt any of the models with time-variation in M or q for management because the estimated variability in M and q was difficult to explain with any known mechanism and management advice differed by nearly an order of magnitude among models. Large differences in management advice underscores the importance of evidence-based approaches to incorporating time-variation in population processes into stock assessment models.

Introduction

Modeling the dynamics of populations often requires making simplifying assumptions to produce tenable models. For example, in fisheries stock assessments, population processes like natural mortality (M), catchability (q), growth, and maturity are often assumed to be time-invariant. However, these assumptions have been shown to be false for some populations. For example, Wilberg et al. (2009) collated published evidence for changes in catchability for harvested marine population, Stawitz et al. (2019) demonstrated considerable variabilty in growth for stocks in the Bering Sea, and Jacobsen and Essington (2018) cataloged changes in natural mortality for forage fish in European waters.

Addressing the potential that other confounded processes are not driving observed changes in a stock is a key concern when estimating time-variation in stock assessment. Similar changes in the observed indices of abundance can be caused by changes in natural mortality, selectivity, growth, or catchability (e.g., Thompson, 1994). For example, a decline in biomass in the terminal year of the time series can arise from increased natural mortality, decreased selectivity or catchability, or decreases in growth. This confounding can lead to the dilemma of 'Kill them or hide them?' when assessing stocks in which there are unexpected declines (Taylor and Methot, 2013). Data to distinguish between mortality events and changes in selectivity are often difficult to procure and, even when the data are possible to collect, may arrive after the point in time at which they are needed to establish management advice (e.g., the next year's survey data). In spite of this, the analyst must make modeling decisions about how to reconcile changes in observed abundance or biomass to provide management advice.

Given the above issues, population dynamics models used to manage harvested natural resources often seek to strike a balance between simplicity and complexity in modeling assumptions that appropriately captures the processes important in determining sustainable harvests (e.g., natural mortality, growth, selectivity, catchability, and maturity) while respecting the information content in the data. In general, three pieces of information are needed in harvest control rules that produce catch recommendations: biomass at the time of the next fishery, a target biomass, and a target fishing mortality. Broadly speaking, these pieces represent the current status of a fishery, the target status for a fishery, and the pathway for how to achieve the target from the current situation. Oftentimes the biomass and fishing mortality targets are based on maximum sustainable yield calculations (e.g., Schaefer, 1954) or proxies thereof (e.g Clark, 1991).

Each of these pieces of information can be influenced by the assumptions about population processes made while developing a population dynamics model. Assuming time-invariance in the population processes that influence estimates of current biomass when those processes are actually varying can produce retrospective patterns and lead to ineffective management advice (e.g., NOAA, 2009). Assuming time-invariance in the population processes that determine the management targets for biomass and fishing mortality may be reasonable if the forcing of these processes is stationary because the targets should represent the mean state of the system. However, if the environmental forcing of these processes is non-stationary (i.e. the mean changes over time), changes in population processes result in changes in management quantities (Thorson et al. 2015) and decisions must be made about the time period used as a reference to calculate management targets (NPFMC, 2007).

Snow crab in the eastern Bering Sea is one such stock that appears to have time-variation in some population processes (e.g., natural mortality in Murphy et al. 2018; catchability in Somerton et al. 2013) and models have been proposed for other stocks of snow crab that include time-varying M (Shibata et al., 2021). The eastern Bering Sea population is at least at risk of non-stationarity in some population processes (particularly those related to ice extent like recruitment; Szuwalski et al., 2020). Consequently, we use it here as a case study to explore the potential to estimate time-variation in confounded population processes.

The snow crab fishery has been managed with individual transferable quotas since 2005 and retains large male crab with carapace width greater than 101 mm (Figure 1; Szuwalski, 2021). Catches were highest in the 1990s, after which the stock biomass declined sharply and the stock was declared overfished in 1999. The stock was declared 'rebuilt' in 2011 once the estimated biomass exceeded the biomass management target. The National Marine Fishery Services (NMFS) summer trawl survey is a key information source in the stock assessment for snow crab. Crab cannot be aged, so true cohorts are not known, however groups of crab

of similar size clearly move through the population and are dubbed 'pseudocohorts'. Large 'pseudocohorts' have appeared in the survey three times since 1982 (Figure 1 and Figure 2).

The most recent large pseudocohort was spawned around 2010 and first selected by the survey gear in 2015 (Szuwalski et al., 2020). By 2018, it was the largest pseudocohort ever observed in the eastern Bering Sea. However, in 2019, it was much smaller than expected given estimated growth, natural mortality, and fishing removals (Figure 2). This strongly implies time-variation in some population process and catchability or natural mortality are key suspects given previous research. However, there are few data to decisively indicate which process is varying over time. No survey was performed in 2020 as a result of the coronavirus pandemic, so the managing body was in the unprecedented position of setting the overfishing level (i.e. the level of catch beyond which overfishing would occur, which is used to set allowable biological catches) in a year with no survey data and in the context of a large drop in the previous year of survey data that could be attributed to changes in either natural mortality or catchability.

Here I present iterations of the assessment method used for snow crab that incorporate time-variation in natural mortality and/or catchability. I then compare the resulting fits to the data, evaluate the character and credibility of the estimated time-variation in q and M, and examine the knock-on effects in other confounded processes like maturity and recruitment. I conclude by presenting the differences in management-related quantities, summarizing the discussion the management body held concerning these models, and discussing potential paths forward.

Methods

The integrated size-structured model currently used (referred to here as the 'status quo' model) was developed by Turnock (2015) following Fournier and Archibald's (1982) methods, with many similarities to Methot (1990). The population dynamics in the status quo model tracks the number of crab of sex s, maturity state m, during year y at width w, $N_{s,m,y,w}$. A terminal molt occurs in which crab move from an immature to a mature state, after which no further molting occurs. The mid-points of the size bins tracked in the model span from 27.5 to 132.5 mm carapace width, with 5 mm size classes. Parameters estimated within the assessment include those associated with recruitment, natural mortality, fishing mortality, selectivity (fishery and survey), survey catchability, and probability of having undergone terminal molt. Weight at size, discard mortality, bycatch mortality, and parameters associated with the variance in growth and proportion of recruitment allocated to size bin were estimated outside of the model or specified. Growth parameters were estimated outside of the model based on measurements of pre-molt and post-molt carapace width for crab captured in the wild just prior to molting. Survey biomass indices and size-composition data, directed fishery biomass and size-composition data, and bycatch biomass and size-composition data were fit to within the assessment. The assessment spans the years 1982 to 2020. See appendix A for a complete description of the population dynamics and model code is available on a github repository linked at the end of this manuscript.

The output of three models are compared to the status quo here: "Vary M", "Vary q", and "Vary both". Each of these models builds on the status quo model and alters the way in which natural mortality ("Vary M"), catchability ("Vary q"), or both processes ("Vary both") are modeled. "Vary M" adds a vector of deviations to the existing estimated parameters for mature natural mortality of both sexes (i.e. a 'dev_vector' in AutoDifferentiating Model Builder; Fournier et al., 2012). The vector of deviations includes an additional parameter for each year included in the assessment (1982-2020) for each sex. "Vary q" estimates a catchability parameter for each year from 1989-2020 for each sex. The estimated vectors of q are bounded by 0.2 and 1.0 based information from experimental net efficiency work for Bering Sea snow crab that suggests catchability should be within this range (Somerton et al., 2013). "Vary both" combines these changes into a single model. A smoothing penalty is added to the negative log likelihood of each model to allow model convergence The smoothing penalty is the squared norm of the second difference of each vector of additionally estimated parameters multiplied by a user-defined weighting factor. The weights for the smoothing penalties were chosen by trial and error, with the aim of making them as small as possible while still avoiding convergence issues.

Retrospective analyses were performed in which the terminal year of data was removed sequentially from the model fitting, then estimated management quantities (e.g., mature male biomass [MMB]) were compared between the model with the complete data set (i.e. 1982-2020) and models with successive 'peels' of the data to identify retrospective patterns. A retrospective pattern is a consistent directional change in assessment estimates of management quantities in a given year when additional years of data are removed from an assessment. Mohn's rho (which computes the average difference between the reference case and the peels over the period of retrospective analysis; Mohn, 1999) was calculated for each model to quantify the retrospective patterns in the estimate of the terminal year of biomass.

The overfishing level (OFL) is the level of catch beyond which overfishing would be deemed to occur in a given year. The OFL is one of the key pieces of management advice that arises from the assessment and serves as the starting point for specifying an 'acceptable biological catch', which constrains removals in the fishery. The OFL was calculated using spawner-per-recruit proxies for biomass and fishing mortality reference points (e.g., Clark, 1991) and a sloped control rule. After fitting the assessment model to the data, the model was projected forward 100 years using the estimated parameters under no exploitation and constant recruitment to determine 'unfished' MMB-per-recruit. For models in which time-variation was implemented, the average of the final seven years of the process was used in projections. The bisection method was used iteratively to identify a fishing mortality that reduced the MMB-per-recruit to 35% of the unfished level (i.e. $F_{35\%}$ and $B_{35\%}$). Calculations of $F_{35\%}$ were made under the assumption that bycatch fishing mortality was equal to the estimated average value.

Calculated values of $F_{35\%}$ and $B_{35\%}$ were used in conjunction with a sloped control rule to adjust the proportion of $F_{35\%}$ that is applied based on the status of the population relative to $B_{35\%}$ (Amendment 24, NMFS; NPFMC, 2007).

$$F_{OFL} = \begin{cases} Bycatch only & if \frac{MMB}{MMB_{35}} \leq 0.25 \\[2ex] \frac{F_{35}(\frac{MMB}{MMB_{35}} - \alpha)}{1-\alpha} & if 0.25 < \frac{MMB}{MMB_{35}} < 1 \\[2ex] F_{35} & if MMB \geq MMB_{35} \end{cases} \tag{1}$$

Where MMB is the projected MMB in the current survey year after fishing at the F_{OFL}, $MMB_{35\%}$ is the equilibrium MMB at the time of mating resulting from fishing at $F_{35\%}$, $F_{35\%}$ is the fishing mortality that reduces the MMB-per-recruit to 35% of unfished levels under constant recruitment, and α determines the slope of the descending limb of the harvest control rule (set to 0.1 here as specified by the NPFMC).

Results

All models produced a positive-definite Hessian and had maximum gradient components less than 0.004, except "Vary both", which had a maximum gradient component of 0.01 for one of the recruitment deviations for males (the rest were < 0.004). The smallest viable smoothing penalties tested were 1 and 10 for natural mortality and catchability (respectively) when they were the only additional time-varying process in the assessment. When both processes were allowed to vary, the smoothness penalty for natural mortality had to be increased to 15 or the model did not converge.

Fits to the data

Models "Vary M", "Vary q", and "Vary both" all fit the survey MMB better than the status quo model (Figure 3 and Table 1), particularly in recent years (Figure 4). The status quo model missed the input confidence intervals of 5 of the last 6 years of survey MMB, but additional estimated time-variation improved these fits. Fits to the catch data, growth data, and all size-composition data sources can be seen in the supplementary materials. In general, the fits were similar across models, except for the survey size-composition data,

4

in which there are several years in which models that allowed for additional time-variation produce much better fits than models that did not. Retrospective patterns in the status quo model (Mohn's rho = 0.34) were improved substantially by allowing additional time-variation in either natural mortality or catchability (Figure 5).

Estimated population processes and derived quantities

Estimated population processes and derived quantities varied among models. The trajectories of the time series of estimated MMB varied strongly among models, and model "Vary both" produced the largest historical estimates of MMB (Figure 6). The shapes of the estimated curve representing the probability of undergoing terminal molt for males were similar among models, but the magnitude of the probabilities varied, most strongly for males in the 70-90 mm carapace width range (Figure 7). 'Vary both' estimated the highest fraction of sub-industry-preferred size (i.e. <101 mm carapace width) crab undergoing terminal molt. Trajectories of estimated fishing mortality were similar across models, but variability within years existed due to differing estimates of population size (Figure 7). Patterns in estimated recruitment were similar for all models, but the relative magnitudes varied (Figure 7).

Estimated natural mortality across models varied widely (Figure 8). Both 'Vary both' and 'Vary M' estimated higher natural mortality in recent years compared to historical estimates, but estimates from 'Vary both' increased sharply starting in 2017 and ended with estimates in 2020 nearly double that of 'Vary M'. Estimates of mean natural mortality were smaller for models in which natural mortality was allowed to vary over time than for those in which natural mortality was constant over time. The shapes of the estimated survey selectivity curves were similar among all models; the largest changes were seen in the catchability coefficient (see supplementary figure 11). Models in which catchability was time-varying estimated average catchabilities similar to the experimentally implied catchability. However, the variability in estimated catchability for these models was large, with estimated values ranging from ~0.2 to ~1 (Figure 8).

Differences in estimated population processes and fits to the data resulted in large differences in management advice (Table 2). The terminal year of estimated MMB varied from 43,290 t to 133,510 t, with models estimating time-varying M returning much lower MMB than those that did not. The corresponding target biomasses ($B_{35\%}$) were also much lower for models incorporating time-varying M because the projections used to calculate $B_{35\%}$ are based on the terminal years of estimated M, which were quite high (Figure 8). An additional effect of increases in estimated M (and changes in the probability of terminally molting; Figure 7) were large increases in $F_{35\%}$ (Table 2). In spite of large differences in the estimated MMB, the management advice from 'Vary q' was only 10% more than 'Vary M' (77.08 vs. 70.88 kt). This similarity in OFLs is a function of the large $F_{35\%}$ and small $B_{35\%}$ from 'Vary M' relative to 'Vary q'. However, if both processes were allowed to vary in the assessment ('Vary both'), the calculated OFL was 81% smaller than the when only one process was allowed to vary. Furthermore, if the smoothness penalty used for 'Vary both' was used in 'Vary M', the calculated OFLs were similar, suggesting relatively unstable estimation of time-varying M.

Discussion

Time-variation in population processes appears to be an issue for snow crab, but data to inform changes in population processes over time are lacking. Allowing M or q to vary over time within the assessment lead to improved fits to the data and reductions in retrospective patterns, but the management advice differed by almost an order of magnitude depending on what processes were allowed to vary and how the processes were constrained within the estimation process. Further, estimates of time-varying natural mortality among assessment variants and other published models fit to the same data (Murphy et al., 2018) were quite different in character (Figure 8). Ultimately, the Crab Plan Team, which is the first level of peer review for the assessments used for crab in the Bering Sea, chose not to endorse models that allowed time-varying q or M for use in management because the estimates of q and M varied more drastically than seemed reasonable and there was no clear mechanism to support the variability.

Improvements in fit and reductions in retrospective patterns with additional model flexibility is an expected outcome and is not unique to either snow crab or size-structured models. Szuwalski et al. (2019) demonstrated a similar outcome in a simulation study based on a cod-like life history with age-structured assessment models. Retrospective patterns were induced by simulating data with time-variation in growth, selectivity, or natural mortality and fitting assessment models that did not model time-variation in those processes. Next, assessment models that allowed either growth, selectivity, or natural mortality were fit to the data. Implementing any of these time-varying processes reduced the magnitude of retrospective patterns, but management advice resulted in over- or under-exploiting the stock if the incorrect process was allowed to vary. One of the main recommendations of the paper was that an understanding of what process is time-varying is recommended before implementation of time-variation in integrated assessments. The lack of a mechanistic understanding for changes in snow crab was the primary reason the candidate models with time-variation in M and/or q were not used in management.

Much effort has been expended exploring methods to estimate time-variation in natural mortality and selectivity in recent years (e.g., Miller and Hyun, 2017; Jiao et al. 2012; Jacobsen et al. 2019). The CAPAM special issue on selectivity (Maunder et al., 2014) presented papers on methodology for using random-effects to estimate time-variation in selectivity (Nielsen and Berg, 2014) and best practices in modeling time-varying selectivity (Martell and Stewart, 2014). Simulation studies have been used to suggest that estimating M at a minimum should be the default in stock assessment (Johnson et al., 2014). More recently, Aldrin et al. (2021) suggested that accounting for variability across time using simple models improves performance over time-invariant M.

In spite of the effort spent estimating time-varying catchability, selectivity, and natural mortality individually, less effort has been spent understanding the feasibility of estimating time-variation in multiple confounded processes within a stock assessment. This is perhaps because the confounding of processes like natural mortality and catchability is recognized as a central challenge in assessment even without considering time-variation (Maunder and Piner, 2014). It has also been shown to be difficult to estimate time variation in a single population process with the data often available for population dynamics modeling, much less two (Johnson et al., 2014, Lee et al., 2011). However, moving forward, the need to estimate time-variation in multiple processes will likely be an increasingly prevalent problem as populations respond to a changing environment (Szuwalski and Hollowed, 2016). Changes in distribution (Pinksy et al., 2013), recruitment (Szuwalski et al., 2015), growth (Audzijonyte et al., 2020), and natural mortality (Audzijonyte et al., 2016) are already being observed in harvested marine populations. If appropriate ways of incorporating multiple processes varying over time into assessment cannot be found, retrospective patterns and erroneous management advice will become increasingly common. This may result in an eroded potency of previously effective management (Hilborn et al., 2021).

Given the outcomes of this analysis, it is not clear if the estimation of time-variation in confounded processes is a problem we can model our way out of with the currently available data for snow crab. Completing these sorts of analyses with a wider range of life histories and data availabilities (both real world and simulated) could be useful to better understand the potential for estimating time-variation in confounded processes in stock assessment. Attention to how to specify smoothness parameters would be useful for practical implementation of time-varying processes in stock assessment given their impact on the analysis here. More longitudinal data are likely needed to directly inform the estimation of time-variation in confounded processes, but the continuous tagging, laboratory, and net efficiency studies that would be needed to provide these data will be expensive. So, cost-benefit analyses aimed at understanding the value of information derived from the needed studies to estimate time-variation in confounded processes should also be performed. These analyses could include management strategy evaluations aimed at understanding the improvement in management outcomes coming from assessments that can estimate time-variation in confounded processes and simpler strategies that do not attempt to do so.

Acknowledgments

Supplementary materials

The github repository including the code used to perform the analysis and additional figures describing the fit of the models to the data can be found at: https://github.com/szuwalski/snow_time_m.

References

Aldrin, M., Aanes, F.L., Tvete, I.F., Aanes, S., Subbey, S. 2021. Caveats with estimating natural mortality rates in stock assessment models using age-aggregated catch data and abundance indices. Fish. Res. doi.org/10.1016/j.fishres.2021.106071

Audzijonyte, A., Richards, S.A., Stuart-Smith, R.D., Pecl, G., Edgar, G.J., Barrett, N.S., Payne, N., Blanchard, J.L. 2020. Fish body sizes change with temperature but not all species shrink with warming. Nat. Eco. Evo. 4(6): 809-814.

Audzijonyte, A., Fulton, E., Haddon, M., Helidoniotis, F., Hobday, A.J., Kuparinen, A., Morrongiello, J., Smith, A.D.M., Upston, J., Waples, R. 2016. Trends and management implications of human-influenced life-history changes in marine ectotherms. Fish Fish. 17(4): 1005-1028.

Clark, W.G. 1991. Groundfish exploitation rates based on life history parameters. Can. J. Fish. Aquat. Sci. 48: 734-750.

Fournier, D.A. and C.P. Archibald. 1982. A general theory for analyzing catch-at-age data. Can. J. Fish. Aquat. Sci. 39:1195-1207.

Fournier, D.A., H.J. Skaug, J. Ancheta, J. Ianelli, A. Magnusson, M.N. Maunder, A. Nielsen, and J. Sibert. 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optim. Methods Softw. 27:233-249.

Hilborn, R., Amoroso, R.O., Anderson, C.M., Baum, J.K., Branch, T.A., Costello, C., De Moor, C.L., Faraj, A., Hively, D., Jensen, O.P., Kurota, H., Little, R.L., Mace, P., McClanahan, T., Melnychuk, M.C., Minto, C., Osio, G.C., Parma, A.M., Pons, M., Segurado, S., Szuwalski, C.S., Wilson, J.R., Ye, Y. 2021. Effective fisheries management instrumental in improving fish stock status. Proc. Natl. Acad. Sci. 117(4): 2218-2224.

Jacobsen, N.S., Essington, T.E. 2018. Natural mortality augments population fluctuations of forage fish. Fish Fish. 19(5): 791-797.

Jacobsen, N.S., Thorson, J.T., Essington, T.E. 2019. Detecting mortality variation to nehance forage fish population assessments. ICES J. Mar. Sci. 76(10): 124-135.

Jiao, Y. Smith. E.P. O'Reilly, R., Orth, D.J. 2012. Modelling non-stationary natural mortality in catch-at-age models. ICES J. Mar. Sci. 69(1): 105-118.

Johnson, K.F., Monnahan, C.C., McGilliard, C.R., Vert-Pre, K.A., Anderson, S.C., Cunningham, C.F., Hurtado-Ferro, F., Licandeao, R.R., Muradian, M.L., Ono, K., Szuwalski, C.S., Valero, J.L., Whitten, A.R., Punt, A.E. 2014. Time-varying natural mortality in fisheries stock assessment models: identifying a default approach. ICES J. Mar. Sci. 72(1): 137-150.

Lee, H.H., Maunder, M.N., Piner, K.R., Methot, R.D. 2011. Estimating natural mortality within a fisheries stock assessment model: An evaluation using simulation analysis based on twleve stock assessment. Fish. Res. 109(1): 89-94.

Martell, S., Stewart, I. 2014. Towards defining good practices for modeling time-varying selectivity. Fish. Res. 158 84-95.

Maunder, M.N. Piner, K.R. 2014. Contemporary fisheries stock assessment :many issues still remain. ICES. J. Mar. Sci. 71(10): 7-18.

Maunder, M.N., Crone, P.R., Valero, J.L., Semmens, B.X. 2014. Selectivity: Theory, estimation, and application in fishery stock assessment models. Fish. Res. 158: 1-4.

Methot, R.D. 1990. Synthesis model: An adaptable framework for analysis of diverse stock assessment data. Int. N. Pac. Fish. Comm. Bull. 50:259-277.

Miller, T.J. and Hyun, S.Y. 2017. Evaluating evidence for alternative natural mortality and process error assumptions using a state-space, age-structured assessment model. Can. J. Fish. Aquat. Sci. 75(5): 691-703.

299 Mohn R. 1999. The retrospective problem in sequential population analysis: an investigation using cod
300 fishery and simulated data, ICES J. Mar. Sci. 56: 473-488.

301 Murphy, J.T., Rugolo, L.J., Turnock, B.J. 2018. Estimation of annual, time-varying natural mortality and
302 survival for Eastern Bering Sea snow crab with state-space population models. Fish. Res. 205: 122-131.

303 Nielsen, A., Berg, C.W. 2014. Estimation of time-varying selectivity in stock assessments using state-space
304 models. Fish. Res. 158: 96-101.

305 NOAA, Report of the retrospective working group. Northeast Fisheries Science Center Reference Documents,
306 09-01, 2009. Woods Hole, MA. National Oceanic and Atmospheric Administration http://www.nefsc.noaa.
307 gov/nefsc/publications/crd/crd0901/crd0901.pdf

308 NPFMC (North Pacific Fishery Management Council). 2007. Environmental Assessment for Amendment
309 24. Overfishing definitions for Bering Sea and Aleutian Islands King and Tanner crab stocks. North Pacific
310 Fishery Management Council,Anchorage, AK, USA..

311 Pinsky, M.L., Worm, B., Fogarty, M.J., Sarmiento, J.L. Levin, S.A. 2013. Marine taxa track local climate
312 velocities. Science. 341(6151): 1239-1242.

313 Schaefer, M.B. 1954. Some aspects of the dynamics of populations important to the management of com-
314 mercial marine fisheries. Bull. Int. Am. Trop. Tuna. Com. 1(2): 27-56.

315 Shibata, Y. Nagao, J. Narimatsu, Y. Eisuke, M., Suzuki, Y., Tokioka, S., Yamada, M., Kakehi, S., Okamura,
316 H. 2021. Estimating the maximum sustainable yield of snow crab off Tohoku, Japan via a state-space stock
317 assessmnt model with time-varying natural mortality. Pop. Ecol. 63: 41-60.

318 Somerton, D.A. Weinberg, K.L., Goodman, S.E. 2013. Catchability of snow crab by the eastern Bering
319 Sea bottom trawl survey estimated using a catch comparison experiment. Can. J. Fish. Aquat. Sci. 70:
320 1699-1708.

321 Stawitz, C.C., Haltuch, M.A., Johnson, K.F. 2019. How does growth mis-specficiation affect management
322 advice derived from an integrated fisheries stock assessment model? Fish. Res. 213: 12-21.

323 Szuwalski, C.S., Vert-Pre, K.A., Punt, A.E., Branch, T.A., Hilborn, R. 2015. Examining common assump-
324 tions about recruitment: a meta-analysis of recruitment dynamics for worldwide marine fisheries. Fish Fish.
325 16(4): 633-648.

326 Szuwalski, C.S. and Hollowed, A.B. 2016. Climate change and non-stationary population processes in fisheries
327 management. ICES J. Mar. Sci. 73(5): 1297-1305.

328 Szuwalski, C.S., Ianelli, J.N., Punt, A.E. 2019. Reducing retrospective patterns in stock assessment and
329 impacts on management performance. ICES J. Mar. Sci. 75(2): 596-609.

330 Szuwalski, C.S., Cheng, W., Foy. R., Hermann. A.J., Hollowed, A.B., Holsman, K., Lee, J., Stockhausen,
331 W., Zheng, J. 2020. Climate change and the future productivity and distribution of crab in the eastern
332 Bering Sea. ICES J. Mar. Sci. 78(2): 502-515.

333 Szuwalski, C.S. 2021. Stock assessment of Eastern Bering Sea snow crab. Stock Assessment and Fishery
334 Evaluation Report for the King and Tanner Crab Fisheries of the Bering Sea and Aleutian Islands Regions.
335 2021 Crab SAFE. North Pacific Fishery Management Council, Anchorage, AK.

336 Taylor, I.G., Methot, R.D. 2013. Hiding or dead? A computationally efficient model of selective fisheries
337 mortality. Fish. Res. 142: 75-85.

338 Then, A.Y., Hoenig, J.M., Hall, G.H., Hewitt, D.A. 2015. Evaluating the predictive performance of empirical
339 estimators of natural mortality rate using information on over 200 fish species. ICES J. Mar. Sci. 72(1):
340 82-92.

341 Thomspon, G.G. 1994. Confounding of gear selectivity and natural mortality rate in cases where the former
342 is a nonmonotone function of age. Can. J. Fish. Aquat. Sci. 51(12): 2654-2664.

343 Thorson, J.T., Monnahan, C.C., Cope, J.M. 2015. The potential impact of time-variation in vital rates on
344 fisheries management targets for marine fishes. Fish. Res. 169: 8-17.

345 Turnock, B.J., 2015. Stock assessment of Eastern Bering Sea snow crab. Stock Assessment and Fishery
346 Evaluation Report for the King and Tanner Crab Fisheries of the Bering Sea and Aleutian Islands Regions.
347 2016 Crab SAFE. North Pacific Fishery Management Council, Anchorage, AK, pp. 167–250.

348 Wilberg, M.J., Thorson, J.T., Linton, B.C., Berkson, J. 2009. Incorporating time-varying catchability into
349 population dynamics stock assessment models. Rev. Fish. Sci. 18(1): 7-24.

Appendix A: Population dynamics model description

Numbers of sex s of shell condition v and maturity state m at width w in the initial year of the assessment, $N_{s,v,m,y=1,w}$, were calculated from an estimated vector of numbers at width w by sex s and maturity state m for males, $\lambda_{s,m,l}$ and numbers at width w by sex s and shell condition v for females (i.e. 2 vectors for each sex were estimated). Estimated vectors of initial numbers at size by maturity for females were calculated by splitting the estimated vectors at size by the observed proportion mature in the first year of the survey. Shell condition is differentiated in the input data for the assessment, but is aggregated over when likelihoods are calculated, so the brief description in the main text does not include shell condition.

$$N_{s,v,m,y=1,w} = \begin{cases} \Omega_{s,w}^{obs}\lambda_{s,1,w} & \text{if v = new; m = mat, s = fem} \\ 1 - \Omega_{s,w}^{obs}\lambda_{s,1,w} & \text{if v = new; m = imat, s = fem} \\ \lambda_{s,2,w} & \text{if v = old; m = mat, s = fem} \\ 0 & \text{if v = old; m = imat} \end{cases} \tag{2}$$

Initial numbers at size for males were all assumed to be new shell.

$$N_{s,v,m,y=1,w} = \begin{cases} \lambda_{s,1,w} & \text{if v = new; m = mat, s = male} \\ \lambda_{s,2,w} & \text{if v = new; m = imat, s = male} \\ 0 & \text{if v = old; m = mat, s = male} \\ 0 & \text{if v = old; m = imat, s = male} \end{cases} \tag{3}$$

The dynamics after the initial year were described by:

$$N_{s,v,m,y+1,w} = \begin{cases} \Omega_{s,w}\kappa_{s,w'}Q_{s,imat,y,w'}X_{s,w',w} & \text{if v = new; m = mat} \\ 1 - \Omega_{s,w}\kappa_{s,w'}Q_{s,imat,y,w'}X_{s,w',w} + Rec_y^{\epsilon}Pr_w & \text{if v = new; m = imat} \\ Q_{s,mat,y,w'} & \text{if v = old; m = mat} \\ (1 - \kappa_{s,w'})Q_{s,imat,y,w'} & \text{if v = old; m = imat} \end{cases} \tag{4}$$

Where $\Omega_{s,w}$ was the probability of having undergone terminal molt at width w for sex s (a freely estimated vector for both males and females constrained by penalties on smoothness), $\kappa_{s,w'}$ was the probability of molting for an immature crab of sex s at width w (set to 1 for all immature crab and zero for all mature crab), and $X_{s,w,w'}$ was the size transition matrix describing the probability of transitioning from size w' to size w for sex s. $Q_{s,m,y,w'}$ was the number of crab of sex s, maturity state m, and width w surviving natural and fishing mortality during year y:

$$Q_{s,m,y,w} = \sum_v N_{s,v,m,y,w}e^{Z_{s,v,m,y,w}} \tag{5}$$

Where $N_{s,v,m,y,w}$ represented the numbers, N, of sex s during year y of shell condition v and maturity state m at width w. $Z_{s,v,m,y,w}$ represented the total mortality experienced by the population and consisted of the sum of instantaneous rates of natural mortality by sex and maturity state, $M_{s,m}$, and fishing mortality, $F_{s,f,y,w}$ from each fishery. Each fishing mortality was subject to selectivity by width w, which varied between sexes s and fisheries f (and by year y if specified). $M_{s,m}$ was specified in the model based on a maximum assumed

age of 20 and Then et al.'s (2015) empirical study on relationships between life history characteristics and natural mortality. A multiplier $\gamma_{natM,m}$ was estimated subject to constraints (this formulation effectively specified a mean and standard deviation for a prior distribution for M).

$$Z_{s,v,m,y,w} = \gamma_{natM,m} M_{s,m} + \sum_f S_{s,f,y,w} F_{s,f,y,w} \tag{6}$$

Selectivities in the directed and bycatch fisheries were estimated logistic functions of size. Different selectivity parameters were estimated for females and males in the directed fisheries ($S_{fem,dir,w}$ and $S_{male,dir,w}$, respectively), a single selectivity for both sexes was estimated for bycatch in the groundfish trawl fishery ($S_{trawl,w}$), and a retention selectivity was estimated for the directed fishery for males ($R_{dir,w}$; all females were discarded).

$$S_{male,dir,w} = \frac{1}{1 + e^{-S_{slope,m,d}(W_w - S_{50,m,d})}}) \tag{7}$$

$$S_{fem,dir,w} = \frac{1}{1 + e^{-S_{slope,f,d}(W_w - S_{50,f,d})}}) \tag{8}$$

$$S_{trawl,w} = \frac{1}{1 + e^{-S_{slope,t}(W_w - S_{50,t})}}) \tag{9}$$

$$R_{dir,w} = \frac{1}{1 + e^{-S_{slope,m,d}(W_w - S_{50,m,d})}}) \tag{10}$$

Where $S_{slope,s,f}$ was the slope of the logistic curve for sex s in fishery f and $S_{50,s,f}$ was the width at 50% selection for sex s in fishery f. Catches for all fisheries were modeled as pulse fisheries in which all catch was removed instantaneously (i.e. no natural mortality occurred during the fishery). Catch in fishery f during year y was calculated as the fraction of the total fishing mortality, $F_{s,f,y,w}$, applied to a given sex s in a fishery f times the biomass removed by all fisheries for that sex.

$$C_{male,dir,y} = \sum_w \sum_v \sum_m w_{male,w} \frac{R_w F_{male,dir,y,w}}{F_{male,dir,y,w} + F_{trawl,y,w}} N_{male,v,m,y,w} e^{-\delta_y M_{s,m}} \left(1 - e^{-(F_{male,dir,y,w} + F_{trawl,y,w})}\right) \tag{11}$$

$$C_{male,tot,y} = \sum_w \sum_v \sum_m w_{male,w} \frac{F_{male,dir,y,w}}{F_{male,dir,y,w} + F_{trawl,y,w}} N_{male,v,m,y,w} e^{-\delta_y M_{s,m}} \left(1 - e^{-(F_{male,dir,y,w} + F_{trawl,y,w})}\right) \tag{12}$$

$$C_{fem,dir,y} = \sum_w \sum_v \sum_m w_{fem,w} \frac{F_{fem,dir,y,w}}{F_{fem,dir,y,w} + F_{trawl,y,w}} N_{fem,v,m,y,w} e^{-\delta_y M_{s,m}} \left(1 - e^{-(F_{fem,dir,y,w} + F_{trawl,y,w})}\right) \tag{13}$$

$$C_{m+f,trawl,y} = \sum_s \sum_w \sum_v \sum_m w_{s,w} N_{s,v,m,y,w} e^{-\delta_y M_{s,m}} \left(1 - e^{-(F_{trawl,y,w})}\right) \tag{14}$$

Where δ_y was the mid point of the fishery (all fisheries were assumed to occur concurrently and the midpoint was based on the directed fishery, which accounts for the vast majority of the fishing mortality) and $w_{s,w}$ was the weight at width w for sex s. Trawl data and discard data were entered into the model with an assumed mortality of 80% and 30%, respectively. Fully-selected fishing mortality parameters for fishery f were estimated as a logged average over a given time period (F_{avg}^{log}) with yearly deviations around that mean ($F_{dev,y}^{log}$).

$$F_{f,y} = e^{(F_{avg,f}^{log} + F_{dev,f,y}^{log})} \tag{15}$$

12

Selectivity for the survey was estimated for 2 eras in the base model: 1982-1988 and 1989-present. Selectivity was assumed to be logistic and separate parameters representing the carapace width at which selection probability equal 50% and 95% ($s_{50,s,e}$ and $s_{95,s,e}$, respectively) were estimated for males and females in the second era (1989-present). Separate catchability coefficients ($q_{s,e}$) were estimated for males and females in all eras.

$$S_{surv,s,w,e} = \frac{q_{s,e}}{1 + e^{-log(19)\frac{W_w - s_{50,s,e}}{s_{95,s,e} - s_{50,s,e}}})} \tag{16}$$

Survey selectivity was informed by experimental surveys during the years 2009 and 2010. A portion of the NMFS summer survey tows were accompanied by an industry vessel using nephrops trawls with an assumed selectivity of 1 for all size classes. To represent the proportion of the population covered by the experiment, a vector was freely estimated for males, S_y^{free} (subject to a scaling parameter), and a logistic curve was estimated for females.

$$S_{ind,s,w,y} = \begin{cases} \frac{q_{ind,s,y}}{1 + e^{-log(19)\frac{L_l - s_{50,s,y}}{s_{95,s,y} - s_{50,s,y}}}}) & \text{if s = female} \\ q_{ind,s,y}S_y^{free} & \text{if s = male} \end{cases} \tag{17}$$

Based on this logic, after identifying the fraction of the crab at length covered by the experimental surveys, the size composition of the NMFS data collected simultaneously with the experimental trawls can be calculated by multiplying the numbers at carapace width 'available' to the experimental trawls by the overall survey selectivity, $S_{surv,s,w,y}$. The predicted numbers at size for the NMFS and industry data from the selectivity experiment were calculated by multiplying the respective selectivities by the survey numbers at size

$$S_{nmfs,s,w,y} = S_{ind,s,w,y}S_{surv,s,w,y} \tag{18}$$

Mature male and female biomass (MMB and FMB, respectively) were fit to in the objective function and were the product of mature numbers at length during year y and the weight at size, $w_{s,w}$:

$$MMB_y = \sum_{l,v} w_{male,w}N_{male,v,mat,y,l} \tag{19}$$

$$FMB_y = \sum_{l,v} w_{fem,w}N_{fem,v,mat,y,l} \tag{20}$$

$$w_{s,l} = \alpha_{wt,s}w_w^{\beta_{wt,s}} \tag{21}$$

Mature biomass can be calculated for different time through out the year, in which case the numbers at size are decremented by the estimated natural mortality. Parameters $\alpha_{wt,s}$ and $\beta_{wt,s}$ were estimated outside of the assessment model and specified in the control file.

Molting and growth occur before the survey. Immature crab were assumed to molt every year with an estimated probability of having undergone terminal molt based on carapace width w (in all the scenarios presented here, the probability of molting was 1 for all immature animals). For crab that molt, the growth increment within the size-transition matrix, $X_{s,w,w'}$, was based on a linear relationship between predicted pre- and post-molt carapace width, ($\hat{W}_{s,w}^{pre}$ and $\hat{W}_{s,w}^{post}$, respectively) and the variability around that relationship was characterized by a discretized and renormalized gamma function, $Y_{s,w,w'}$.

$$X_{s,w,w'} = \frac{Y_{s,w,w'}}{\sum_{w'} Y_{s,w,w'}} \tag{22}$$

$$Y_{s,w,w'} = (\Delta_{w,w'})^{\frac{\hat{L_{s,w}} - (\bar{W}_w - 2.5)}{\beta_s}} \tag{23}$$

$$\hat{L}^{post}_{s,w} = \alpha_s + \beta_{s,1} hat W^{pre}_{s,w} \tag{24}$$

$$\Delta_{w,w'} = \bar{L}_{w'} + 2.5 - W_w \tag{25}$$

$\hat{W}^{post,1}_{s,l}$ was the predicted post-molt carapace width and $\Delta_{w,w'}$ is the molt increment at size.

An average recruitment for the assessment period (1982-present) and yearly deviations around this average were estimated for both males and females. This modeling assumption was implemented because there appear to be differences between sexes in recruitment to the model and not allowing for these differences resulted in large retrospective patterns. Each year's estimated recruitment was allocated to size bins based on a discretized and renormalized gamma function with parameters specified in the control file.

$$Rec_{y,s} = e^{(Rec_{avg,s} + Rec_{dev,y,s})} \tag{26}$$

$$Pr_w = \frac{(\Delta_{1,w})^{\alpha_{rec}/\beta_{rec}} e^{-\Delta_{1,w'}/\beta_{rec}}}{\sum_{w'} (\Delta_{1,w'})^{\alpha_{rec}/\beta_{rec}} e^{(-\Delta_{1,w'}/\beta_{rec})}} \tag{27}$$

For models in which separate vectors of recruitment deviations were estimated for males and females, a separate average recruitment was also estimated (in log space). Each vector of deviations was also subject to a smoothing penalty, but were not linked directly in any way (e.g., priors on the ratio of estimated male to female average recruitment).

Three general types of likelihood components were used to fit to the available data. Multinomial likelihoods were used for size-composition data, log-normal likelihoods were used for indices of abundance data, and normal likelihoods were used for catch data, growth data, priors, and penalties. Multinomial likelihoods were implemented in the form:

$$L_x = \lambda_x \sum_y N^{eff}_{x,y} \sum_l p^{obs}_{x,y,l} ln(\hat{p}_{x,y,l}/p^{obs}_{x,y,l}) \tag{28}$$

L_x was the likelihood associated with data component x, where λ_x represented an optional additional weighting factor for the likelihood, $N^{eff}_{x,y}$ was the effective 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.

Log normal likelihoods were implemented in the form:

$$L_x = \lambda_x \sum_y \frac{(ln(\hat{I}_{x,y}) - ln(I_{x,y}))^2}{2(ln(CV^2_{x,y} + 1))} \tag{29}$$

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, $I_{x,y}$ was the observed value of quantity I from data component x during year y and $CV_{x,y}$ was the coefficient of variation for data component x during year y.

Normal likelihoods were implemented in the form:

$$L_x = \lambda_x \sum_y (\hat{I}_{x,y} - I_{x,y})^2 \tag{30}$$

L_x was the contribution to the objective function of data component x, λ_x was represents the weight applied to the data component (and can be translated to a standard deviation), $\hat{I}_{x,y}$ was the predicted value of

quantity I from data component x during year y, $I_{x,y}$ was the observed value of quantity I from data component x during year y.

Smoothing penalties were also placed on some estimated vectors of parameters in the form of normal likelihoods on the second differences of the vector with a user-specified weight. Weightings of all likelihoods and penalties can be found in the .CTL file in the github repository associated with this paper noted in the acknowledgements.

Table 1: Contribution to the objective function by individual likelihood component by modeling scenario.

Likelihood component	Status quo	Vary q	Vary M	Vary both
Recruitment deviations	71.12	70.7	62.97	69.83
Initial numbers old shell males small length bins	2.83	2.86	2.81	2.59
ret fishery length	194.07	189.63	182.27	180.54
total fish length (ret + disc)	560.36	554.67	546.02	536.71
female fish length	127.32	128.3	129.98	126.79
survey length	2298.23	2275.79	2208.05	2228.97
trawl length	169.34	164.05	151.31	165.36
2009 BSFRF length	-45.58	-45.9	-46.92	-44.82
2009 NMFS study area length	-36.62	-37	-37.86	-35.78
M multiplier prior	34.55	40.78	11.44	13.14
maturity smooth	45.6	42.6	29.55	34.71
growth males	0	0	0	0
growth females	0	0	0	0
2009 BSFRF biomass	0.49	0.03	0.09	0.15
2009 NMFS study area biomass	0.34	0.04	0.06	0.08
cpue q	0.4	0.48	0.19	0.56
retained catch	6.02	3.91	1.21	1.69
discard catch	106.68	116.94	41.42	46.54
trawl catch	11.64	5.62	7.85	5.53
female discard catch	4	4.67	5.71	7.03
survey biomass	215.85	126.53	79.96	77.08
F penalty	23.97	21.48	23.83	22.32
2010 BSFRF Biomass	5.53	2.93	3.4	1.56
2010 NMFS Biomass	6.42	1.85	3.55	4.14
Extra weight survey lengths first year	273.98	274.43	267.88	265.7
2010 BSFRF length	-24.2	-23.6	-22.6	-23.08
2010 NMFS length	-30.65	-29.33	-30.84	-31.08
smooth selectivity	1.22	1.3	1.26	1.22
smooth female selectivity	0	0	0	0
init nos smooth constraint	32.25	32.84	31.03	31.17
Total	4055.16	3926.6	3653.62	3688.65

Table 2: Reported management quantities for each scenario considered. Reported MMB is for the final year, natural mortality (M) is the average over the entire time series for mature males, and average recruitment (avg_rec) is for males.

Model	MMB	B35	F35	FOFL	OFL	M	avg_rec
Status quo	133.51	121.47	1.23	1.23	88.90	0.29	103.91
Vary q	121.61	137.56	1.94	1.94	77.08	0.30	132.86
Vary M	43.29	17.85	6.29	6.29	70.88	0.59	152.61
Vary both	92.20	28.06	12.46	1.86	14.72	0.58	241.96
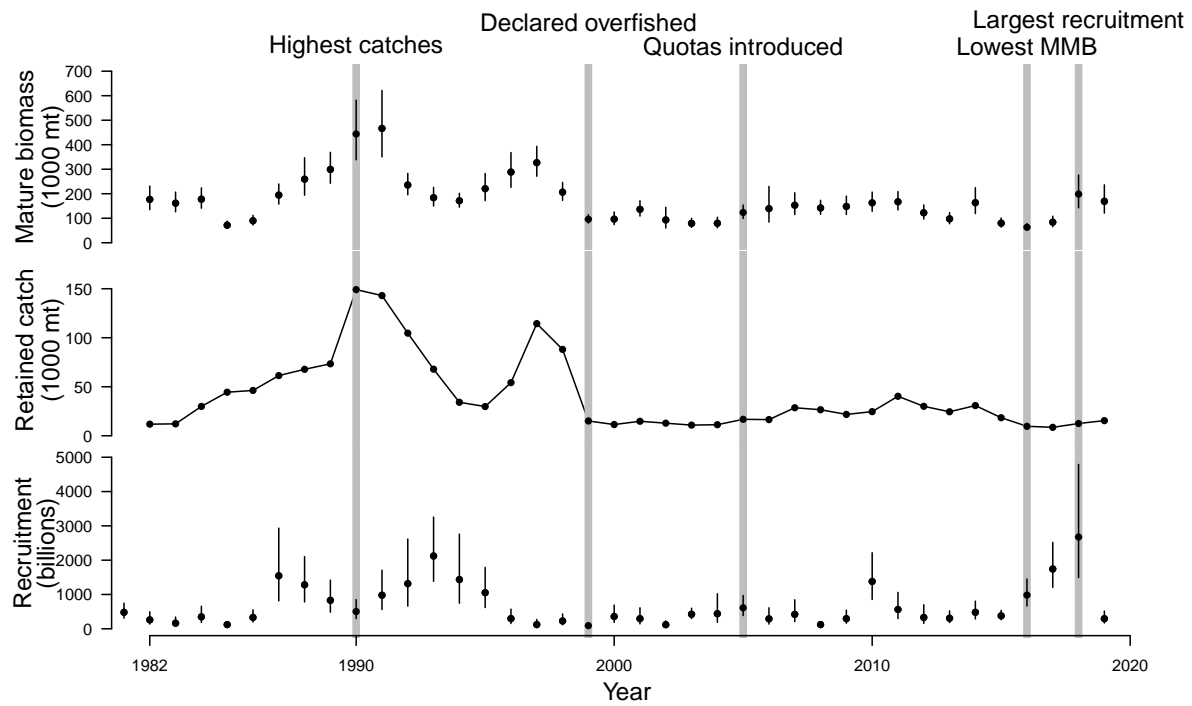
Figure 1: Graphical history of the snow crab fishery in the eastern Bering Sea. Mature biomass are the observed values inthe survey, retained catch are reported from fish ticket data, and recruitment is the number of crab sized 40 to 55 mm carapace width observed in the survey. Bars on MMB and recruitment are 95% confidence intervals.
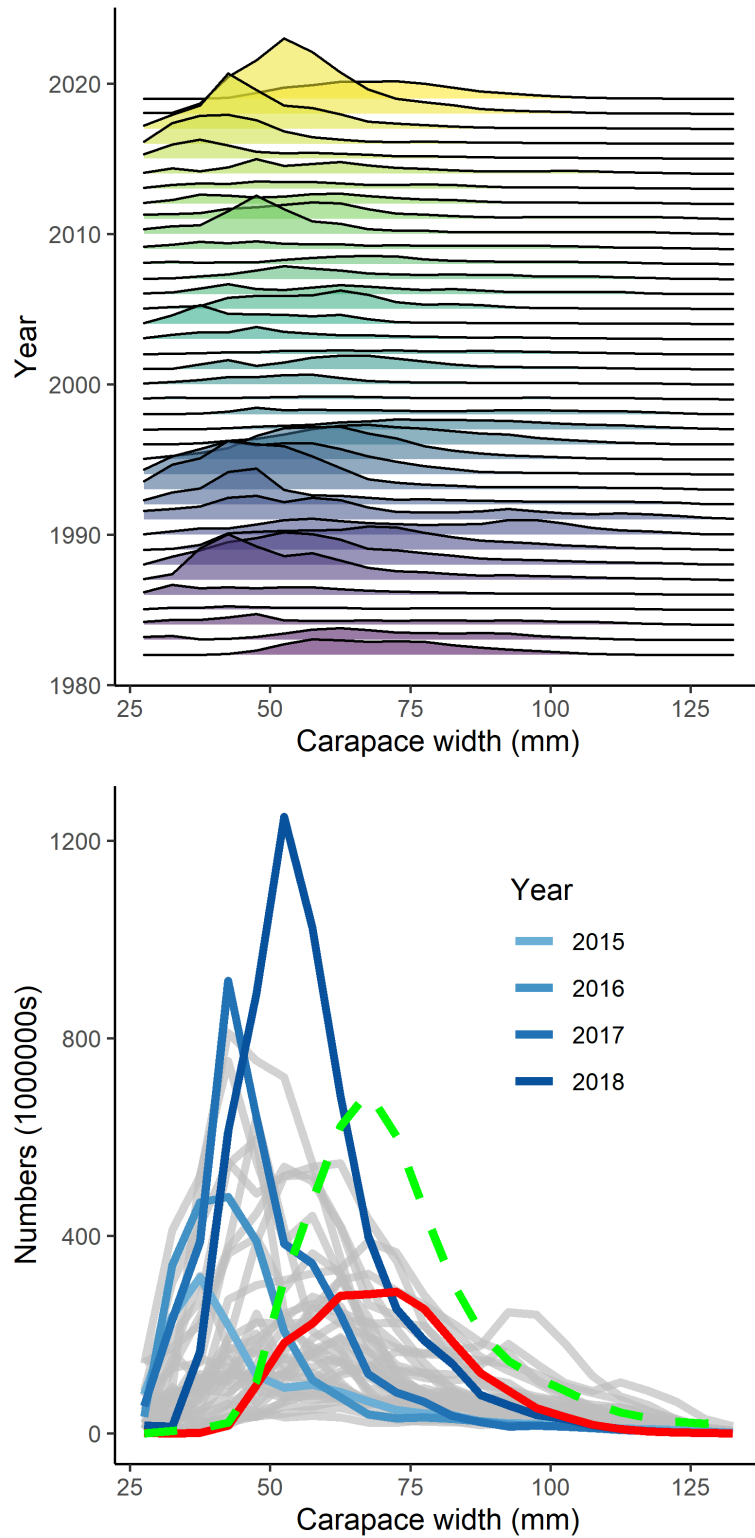
Figure 2: Numbers at length from the survey. Top figure shows the relative numbers at length over time in which the height of the polygon for a given year represents the number of crab observed in that size class. Bottom figure shows the same data, but overlaid. Highlighted years show the recruitment of the 2010 cohort to the survey gear in 2015 and the subsequent reduction of the cohort in 2019 (red line). Dashed green line represents the approximate expected numbers at length given an application of growth, fishery removals, and natural mortality to the numbers at length from 2018.
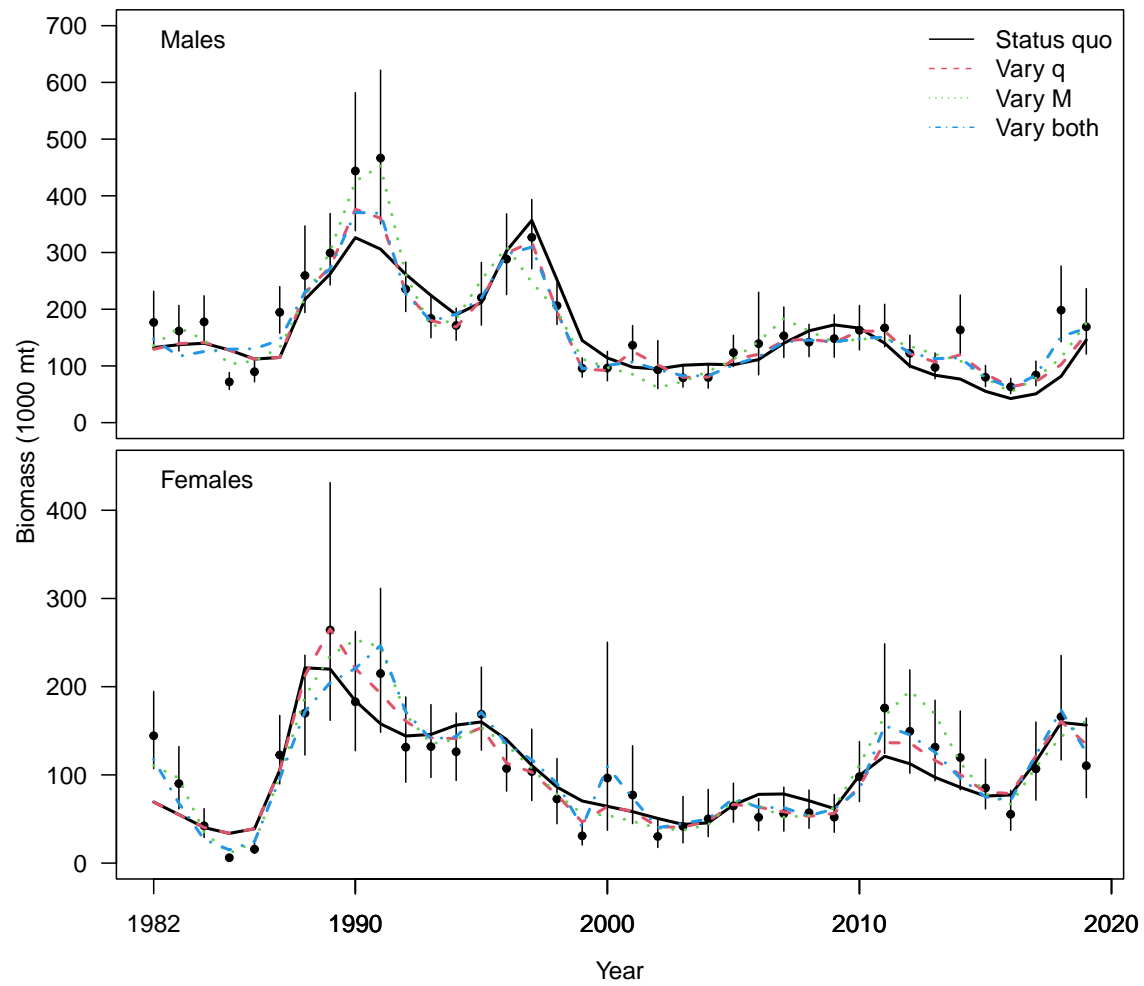
Figure 3: Model fits to the observed mature male biomass at survey with associated 95% confidence intervals (bars).
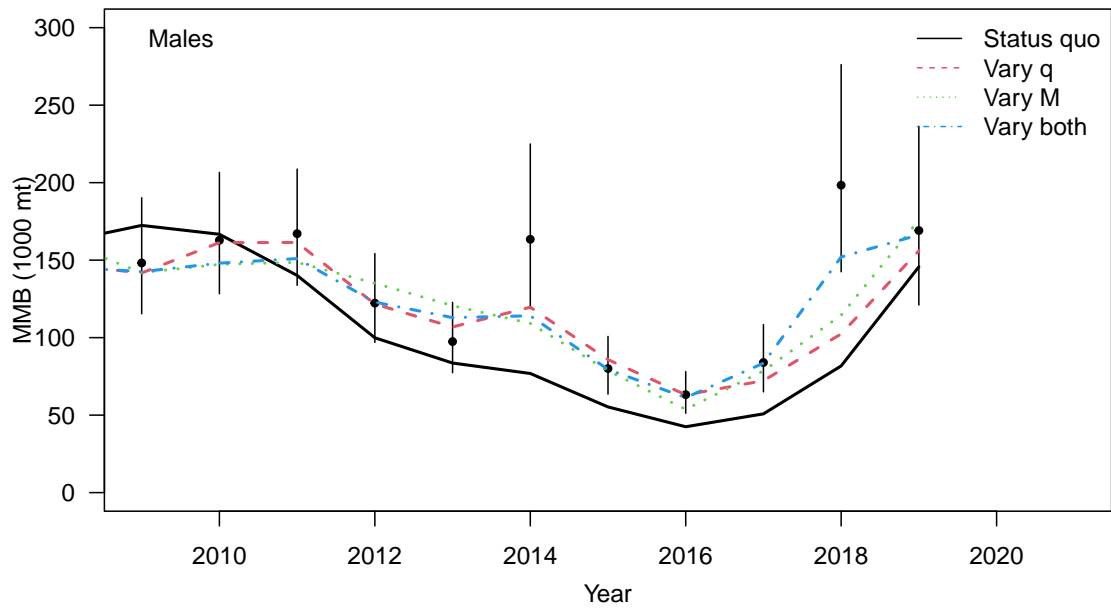
Figure 4: Model fits to the observed mature male biomass at survey 2009-present, note a change in the y-axis from figure 3. Vertical bars are 95% confidence intervals.
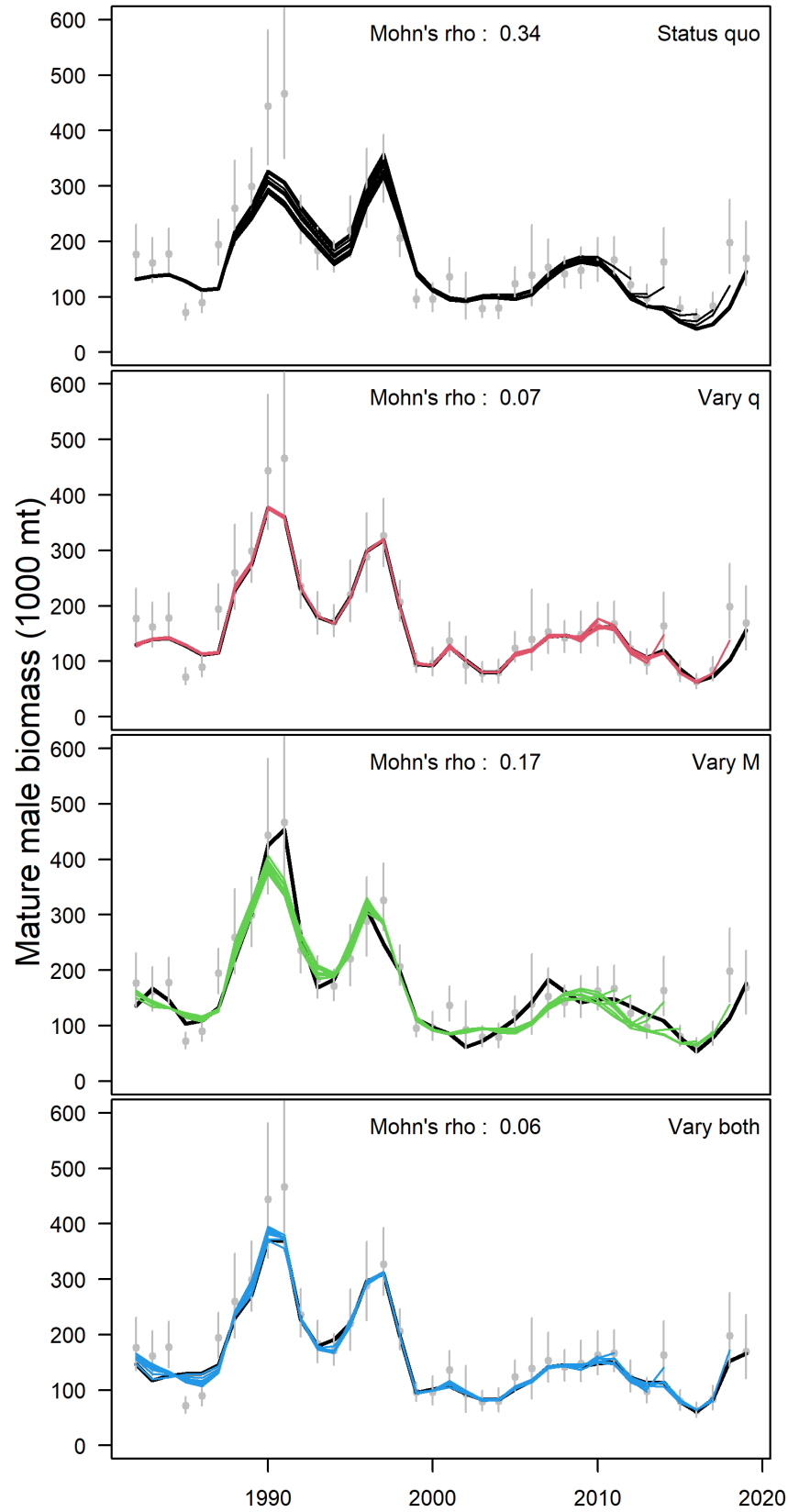
Figure 5: Retrospective patterns in mature male biomass (grey points with 95 percent confidence intervals) for all models in the analysis. Mohn's rho represents the average deviation of estimated MMB of each peel from the reference peel (black lines).
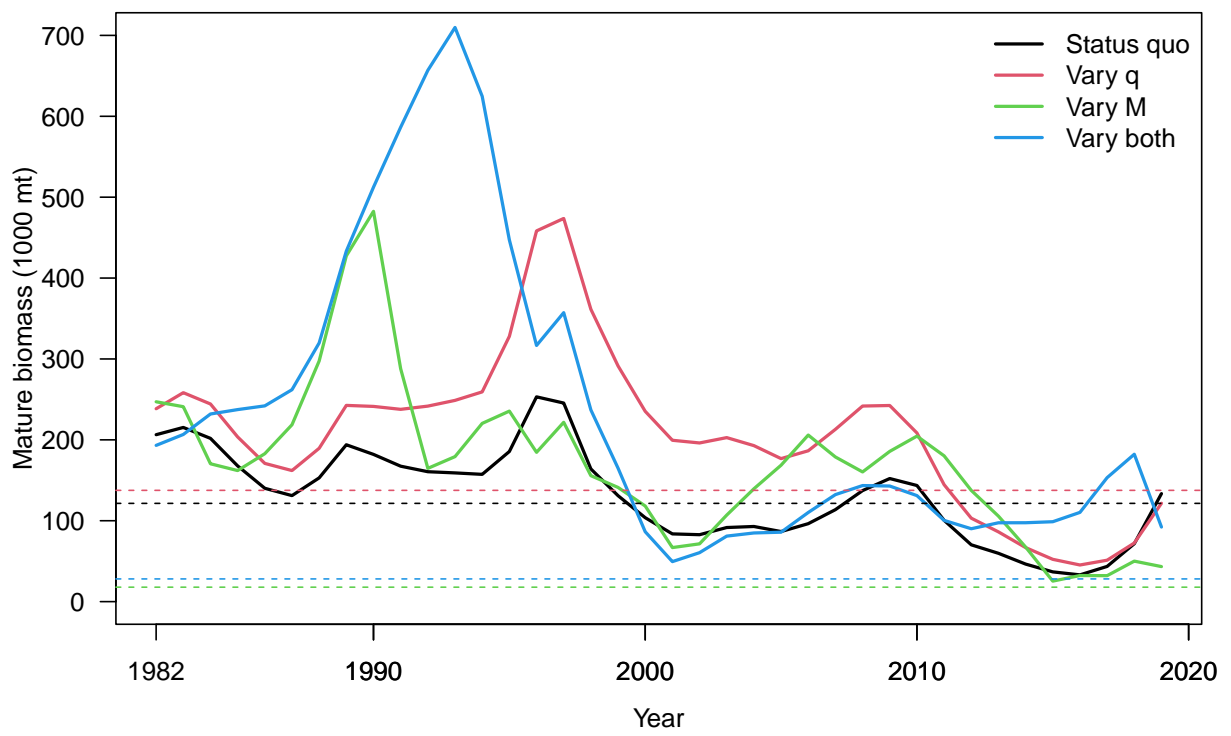
Figure 6: Predicted mature biomass at mating time. Dotted horizontal lines are target biomasses.
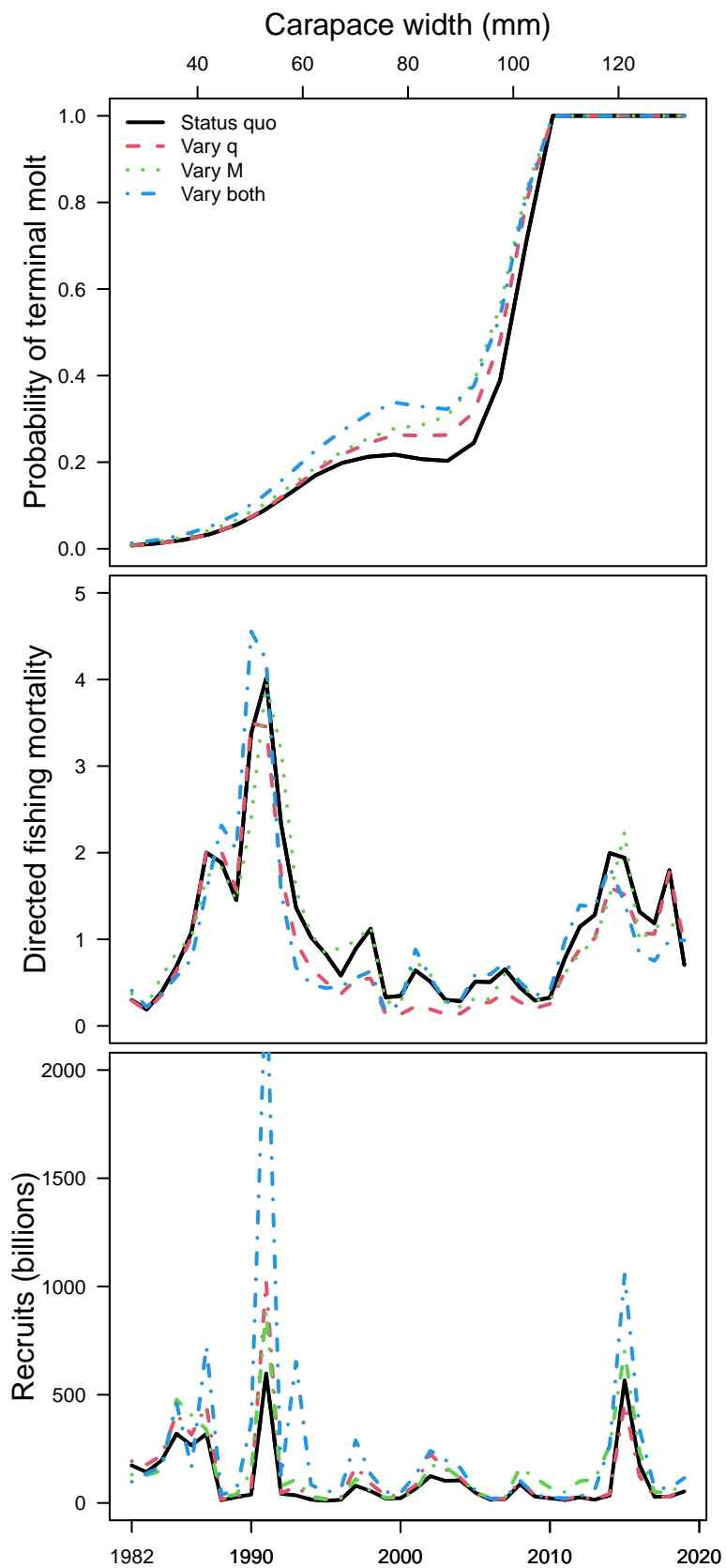
Figure 7: Estimated probability of having undergone terminal molt, directed fishing mortality, and recruitment.
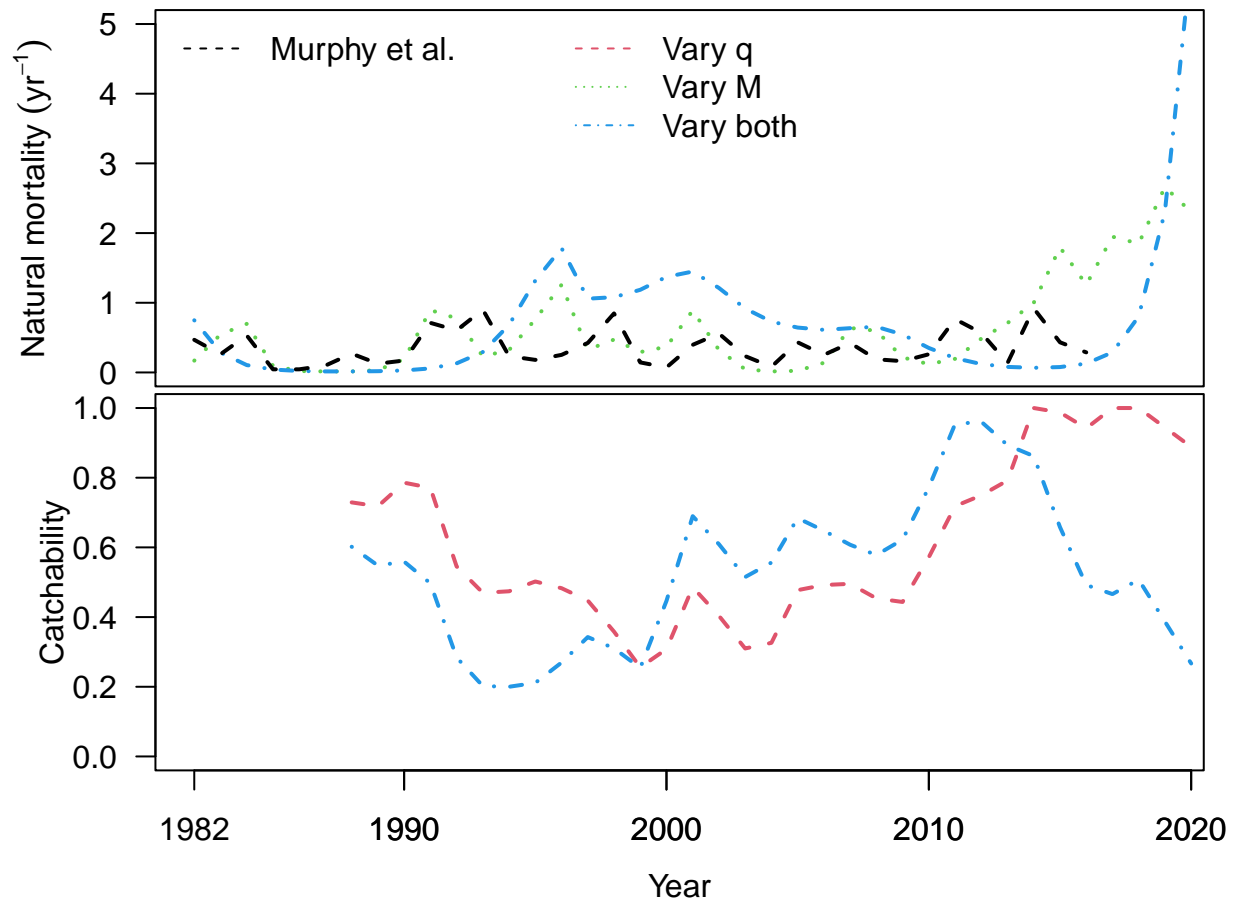
Figure 8: Estimated instantaneous rate of time-varying natural mortality and estimated catchability for mature males.