Summary Report of the Chesapeake Bay Blue Crab Stock Assessment Peer Review

April 13 - 15, 2026

Annapollis, MD

Report prepared by Panel Members:

Dan Hennen (Chair), NEFSC

Simon de Lestrang, DPIRD, Australia, CIE reviewer

Nick Caputi, DPIRD, Australia, CIE reviewer

Yong Chen, Stony Brook University, NY, CIE reviewer

Introduction

The NOAA Chesapeake Bay Office (CBO) convened a review of the Benchmark Quantitative Fishery Stock Assessment for Blue Crab in the Chesapeake Bay (blue crab stock assessment). The previous blue crab stock assessment was completed in 2011.

The review was conducted by three members of the Center for Independent Experts (CIE) and one member of the Chesapeake Bay Stock Assessment Committee (CBSAC). The review panel (panel) members were Nick Caputi from Western Australian Fisheries and Marine Research Laboratories, Department of Primary Industries and Regional Development, Australia (CIE), Yong Chen from the State University of New York at Stony Brook (CIE), Simon De Lestang from the Sustainability and Biosecurity Department of Primary Industries and Regional Development, Australia (CIE), and Dan Hennen from the National Marine Fisheries Service (CBSAC). The panel was chaired by Dan Hennen and assisted by Bruce Vogt (CBO), Christina Garvey (CRC) and Nick Staten (CRC).

The Assessment Team (AT) was led by Mike Wilberg from the University of Maryland Center for Environmental Science. Other AT members included: Maya Drzewicki[1], Gina M. Ralph[2], Dong Liang[1], Madison Sholes[1], Robert Aguilar[3], Amanda Bevans[4], Ingrid Braun-Ricks[5], Mandy Bromilow[6], Glenn R. Davis[6], Alexa Galvan[7], Thomas Ihde[4], Robert Latour[2], Romuald. N. Lipcius[2], Brooke Lowman[7], Marvin M. Mace III[6], Thomas J. Miller[1], Matthew B. Ogburn[3], Gabrielle Saluta[2], Alexandra K. Schneider[8], Alexei Sharov[6], and Troy D. Tuckey[2].

The blue crab stock assessment report, 13 working papers, base-run modeling results, R code for the Assessment model and several background papers were made available to the panel ahead of the review. The review meeting was held on April 13-15, in Annapolis, MD with a remote option so that the Australian reviewers could participate. AT members presented their work on a pre-defined set of terms of reference as well as additional analyses requested by the panel during the meeting. The meeting agenda is provided as Appendix 1.

The AT provided clear descriptions of their work on each of the TOR and were extremely responsive to panel requests for extra analyses throughout the meeting. The panel commends their excellent work.

The panel agreed that all 11 of the TOR were at least partially met and that this assessment represents a large step forward from the 2011 benchmark assessment. The current assessment represents the best science available and should be used for management advice going forward.

The panel's review of each of the 11 TOR is provided below.

---

[1] University of Maryland Center for Environmental Science, Chesapeake Biological Laboratory, Solomons, MD

[2] Virginia Institute of Marine Science, William & Mary, Gloucester Point, VA

[3] Smithsonian Environmental Research Center, 647 Contees Wharf Road, Edgewater, MD

[4] Morgan State University, Patuxent Environmental & Aquatic Research Laboratory (PEARL), St. Leonard, MD

[5] Potomac River Fisheries Commission, Colonial Beach, VA

[6] Maryland Department of Natural Resources, Fishing and Boating Services, Tawes Administration Building, Annapolis, MD

[7] Virginia Marine Resources Commission, Fort Monroe, VA

[8] University of Maine, School of Marine Sciences, Portland, ME

Evaluation of the terms of reference

TOR 1: Critically review and estimate life history parameters and vital rates of blue crab in the Chesapeake Bay that are relevant to the stock assessment. In particular, the assessment should evaluate the extent and scale of interannual variation in life history parameters and vital rates of blue crab in the Chesapeake Bay.

Panel Decision: Partially met.

The AT undertook a very thorough review of the extensive literature that exists on life history of blue crabs in Chesapeake Bay and beyond. The genetic status of blue crabs in Chesapeake relative to crabs in other regions is well studied. The evidence supported the assumption that the blue crab population in Chesapeake Bay comprises a unit stock for assessment purposes. The current assessment has conducted new analyses to estimate length-weight relationships, natural mortality, overwinter mortality (OWM), size at recruitment and female maturity and growth.

Ogburn et al. (2026, Working Paper 7) used data from a long-term (2004-2023), monthly trawl survey of blue crabs in the Rhode River, Maryland to evaluate 1) the timing of the molt to maturity, 2) the growth increment at the molt to maturity, and 3) variability in the operational sex ratio. The proportion of females maturing varied among months with the abundance of pre-pubertal females peaking in August. It would be useful if other blue crab surveys also began identifying pre-pubertal females to understand the spatial and temporal variation in female maturation and growth. Information on pre-pubertal female abundance could also be considered for incorporation in the sex- and length-structured stock assessment model (SLAM) and this could provide an indication of mature female abundance a year ahead as the annual survival of mature females is estimated to be about 10% (Lipcius et al. 2026a).

There has been increased protection of mature females since the management changes after 2008 and the mature female abundance has increased, however the male abundance has declined. The current status of the operational sex ratio (OSR) is estimated to be satisfactory (Ogburn et al. 2026, Working Paper 7) but it is important that the OSR continues to be monitored especially as the male abundance has been below the target reference level for the last four years.

The mean annual survival rate of mature female crabs was 0.10 ± 0.01 (yr^{-1}; mean ± SE), and the mean recovery rates were 0.22 ± 0.01 during 2002-2008 and 0.08 ± 0.004 during 2009-2017 (Lipcius et al. 2026a). The decline in recovery rate is likely to reflect a decrease in fishing effort and supports the decrease in fishing mortality on mature females estimated by SLAM.

Interannual variation in life history parameters has been assessed for some life history parameters (e.g., overwinter mortality), but the interannual and spatial variation in some life history rates (e.g. growth, maturity) were not fully explored.

Several approaches to estimating growth were tried including temperature-dependent growth and a von Bertalanffy growth model, but the Gompertz model was finally selected. A variable b parameter fit the size distribution of the survey crab length data best. However, this was done in an early stage of the model development. Further assessment of the growth model is indicated in the base model with consideration of spatial and temporal variation in growth as there is a north-south gradient in growth rate.

The Stock Assessment Report states (Page 34) that 'observed variability in winter temperatures could vary the timing of recruitment to the fishery by up to 10%' which would affect the size distribution of the recruits in the model. It is important to continue monitoring the timing of recruitment as this is probably sensitive to temperature changes which are influenced by climate change. The SLAM model currently assumes a fixed (log normal) size distribution and pulse recruitment in February, which may be inconsistent with the molting events observed in the field. The impacts of recruitment timing on the stock assessment should be carefully evaluated.

About a third of the crab natural mortality is believed to be due to cannibalism. As there are large variations in annual abundance, there could be marked variations in annual natural mortality due to a density-dependent mortality rate. This should be assessed and examined in SLAM as a sensitivity test.

The environmental drivers affecting recruitment are being examined in a separate study and were not reported at this review. Understanding the drivers of recruitment variation (egg production and environmental conditions) has become especially important with the recent recruitment decline at a time when the mature female abundance is above target level. Historically there have been declines in recruitment when the abundance of mature females has been low and there is some evidence that overfishing has occurred at times. The environmental assessment on recruitment combined with the information available on mature female abundance should enable an assessment of the stock-recruitment-environmental relationship.

TOR 2: Describe and quantify patterns in fishery-independent surveys to develop indices of abundance and characterize the size composition of the population. Analyses should include: (1) A comprehensive evaluation of the utility of fishery-independent surveys to inform the stock assessment; (2) Consideration of index standardization which may include effects of environmental and abiotic factors on survey catches; and (3) Characterization of uncertainty in indices of abundance.

Panel Decision: Partially met.

The AT treatment of fishery-independent data represents a substantial advance over the 2011 benchmark. Eight long-term surveys were evaluated for their utility, with clear rationale for their subsequent inclusion or exclusion from the assessment. The adoption of VAST spatiotemporal models for random-design surveys and GAMLSS for fixed-site surveys is appropriate and well-implemented, with rigorous model selection across spatial/spatiotemporal random field configurations and distributional assumptions. Vessel calibration work for both the VIMSTrawl Survey and ChesMMAP was thorough, supported by years of paired-tow experiments. The development of monthly indices to match the SLAM model's monthly time step is a clear improvement over the annual indices used previously, and the coherence analysis using Kendall rank correlations provides useful transparency about agreement among data sources.

There are no examinations of the spatial coherence of the survey data even though the surveys differ markedly in spatial coverage. The WDS spans the entire Chesapeake Bay, while the VIMS Trawl Survey covers only the area south of the Rappahannock River, ChesMMAP samples the mainstem, the Maryland Trawl Surveys are restricted mostly to eastern shore tributaries, and the PEARL Pot Survey monitors just a few sites near Calvert Cliffs. The relatively low among-survey coherence (Kendall τ ranging from -0.04 to 0.46) is most likely a consequence of these surveys sampling different portions of the Bay that support different components of the population, which may show spatial variability in stock dynamics. Before attempting to reconcile these indices within a single spatially-aggregated model, the assessment team should examine whether the size composition data and population demographics from each survey are consistent with one another (taking into account survey selectivity and spatial coverage). Where size compositions/demographics are similar, the surveys are likely sampling similar segments of the population and can reasonably be treated as observing the same modelled quantity. Where they diverge, combining them under a single selectivity function introduces a mismatch between what the survey observes and what the model predicts. Breaking the WDS into spatial sub-regions and treating them as separate fleets in the model, each with its own selectivity, would likely improve model fit and better represent the known spatial heterogeneity in blue crab population structure across the Bay, without requiring a more complex model structure (multiple areas).

A related concern is the inconsistency between how abundance indices and size composition data were treated. The indices were carefully standardised using spatiotemporal or GAM-based approaches that account for spatial, temporal, and vessel effects, yet the size composition data fed into the model were not standardised in the same way. This creates a potential mismatch in the signals being provided to the assessment model. The index says one thing about trends in abundance after removing confounding factors, but the raw size composition data may carry spatial or temporal biases that have not been similarly addressed. There are two approaches to resolving this: either standardise the length-frequency data using comparable model-based methods (which is quite computer intensive and difficult), or subset the spatial coverage of each survey to identify regions where size compositions are consistent, thereby ensuring that the selectivity assumed in the model is appropriate for the spatial domain actually being indexed. The latter approach has the additional advantage of clarifying the appropriate spatial footprint over which each survey should be truncated to inform the model.

The TOR explicitly asks for consideration of environmental and abiotic factors in index standardisation. While the VAST framework provides terms for density and catchability covariates, in practice, only vessel effects and area-swept offsets were employed. Temperature and salinity data are collected at most survey stations and are known to affect blue crab availability. Productivity of the system, based on temporal and spatial proxies of nutrient load, are also available. This is particularly important for the WDS, where water temperature directly determines whether crabs are torpid and available to the dredge gear. The spatiotemporal random fields in VAST will capture some environmental variation implicitly, but an explicit evaluation of whether including environmental covariates improved index standardisation was not reported. Additionally, the report would benefit from showing the incremental impact of each standardisation component on the resulting index. Presenting the index sequentially: raw, then with spatial effects, then with vessel effects, then with spatiotemporal effects allows the reader to assess how much each factor contributes and whether any single adjustment is driving the overall trend. Some plots associated with the standardisation were provided upon request by the AT, but their inclusion into the report would aid transparency.

Further, the censoring of recruits from summer–winter surveys using size cutoffs (< 35 mm in August, < 65 mm September–January) impacts the index magnitude and appears to have been determined by visual inspection of length-frequency distributions rather than through formal analysis. Since these cutoffs directly affect both the indices and the size compositions entering the model, and since the model's recruitment estimates depend on what is excluded, a sensitivity analysis on the cutoff values would strengthen confidence in the approach. It is also likely that the delineation between cohorts varies temporally (between years) with changes in recruitment time and growth rates. More representative delineation could be determined through the fitting of year-specific mixture models.

TOR 3: Describe and quantify patterns in catch, effort, and CPUE. Analyses should include: (1) Estimation of catch and effort for each jurisdiction; (2) Evaluation of the utility of a commercial CPUE index in the assessment; (3) Examination of the impacts of reporting changes and trends in CPUE; (4) Evaluation and quantification of bycatch and/or discard mortality, and recreational harvest using available data from the jurisdictions; and (5) Characterization of uncertainty in the data.

Panel Decision: Fully met.

Monthly catch and effort were available for each management jurisdiction (Maryland, Virginia, and the Potomac River) for all commercial fishers during 1994-2023. There has been mandatory reporting of commercial landings available since 1994 for all jurisdictions which resulted in improved catch data in this assessment. Catches were examined for six fleets using eight gear or condition types: Maryland Pot, Maryland Trotline, Virginia Pot, Virginia Dredge, Potomac River, and peeler.

An attempt was made to account for recreational harvest and dead discards of blue crabs in the commercial fishery. For example, since the last assessment, a tagging study by Semmler et al. (2021) estimated the recreational harvest of male blue crabs in Maryland. The dead discards are based on the number of dead crabs recorded in pots. However, there is no accounting for incidental mortality of undersize and other crabs returned to the sea alive.

The sex ratios and size compositions of the catch for each fleet were estimated from commercial monitoring and relevant survey data. The proportions-at-length and sex of the fishery catch in Maryland, the Potomac River, and the Bay-wide peeler/soft shell fishery, were informed by data from the Maryland Sentinel Fishery Monitoring Program. Through this program, biological data were collected from commercially caught blue crabs by state biologists and commercial fishers starting in 2002 (Fegley et al. 2006). However, the length compositions for the Virginia pot fishery were based on data from a buyer sampling program that occurred in 2016-2017. More regular monitoring of the sex and length composition of the Virginia landings data is required if this information is to be used in the SLAM.

Estimates of uncertainty in the catch estimates were mainly associated with the estimates of dead discards and recreational harvest.

The analysis of 50+ years of data from the PEARL Pot Survey suggested a significant decline in average size for both males and females within the survey area, with much of the decline in male CW occurring prior to 1990. This appears to have stabilised during the period of assessment 1994-2023. The long-term assessment of mean size trends can be an important indicator of fishing effort so it is important that this continues to be monitored.

Harvest controls in the form of daily bushel limits were instituted in the 2010s and are now in place throughout Chesapeake Bay. Some fishers (~25% in Virginia) are recording daily catches at the bushel limits.

While catch reporting is mandatory, there is evidence that there may be some invalid reporting by some fishers. There is some validation of harvest reports but the error rate is not recorded. It is important to have some validation of the commercial landings with a recording of the error rate and the nature of the error found.

Commercial fishing effort (and hence catch per unit effort, CPUE) was evaluated but was not considered sufficiently reliable to be considered in SLAM. The availability of fishery-independent surveys provides more reliable estimates of abundance trends than would be provided by commercial CPUE. However, understanding the trends in fishing effort, particularly in the pot fisheries which take most of the catch (e.g. 80-97% of annual harvest in Virginia), would provide information relevant to trends in fishing effort and could help interpret model estimates of fishing mortality. Information on the number of trips per year was presented at the review but additional data on the number of traps used would be useful.

TOR 4: Evaluate the feasibility of, and if possible, implement blue crab stock assessment models that operate on: sub-annual time steps and/or at spatial resolutions lower than that of the entire Chesapeake Bay to better represent population dynamics.

Panel Decision: Partially met

The development of the Sex and Length Structured Assessment Model (SLAM) is a substantial improvement for the Chesapeake Bay blue crab stock assessment, moving from the annual age-based catch survey model used in 2011 to a monthly, length- and sex-structured framework implemented in RTMB. The model code is clear and well structured, and the process of tracking males, immature females, and mature females through 1-cm length bins, applying length-based natural mortality, overwinter mortality, and fleet-specific fishery selectivity is appropriate for this species. The decision to use RTMB provides automatic differentiation and access to random effects. The model converged with a small maximum gradient and an invertible Hessian, and the jitter analyses, while revealing some instability associated with PEARL catchability and ChesMMAP selectivity, supported the robustness of the chosen base model. However, the report's description of the model development process was incomplete and did not fully reflect the sequence of decisions and iterations that led to the final model structure, as described during the review process.

A concern with the current parameterisation is the estimation of separate fishing mortality intensities for every month × sex × fleet combination. While sex-specific fishing mortality is biologically plausible given different regulations and spatial distributions, estimating fully independent F values by sex means that any discrepancy between observed and predicted sex ratios in the catch can be absorbed into F rather than being attributed to selectivity or population structure. This can mask real conflicts in the data. If male and female F estimates are highly correlated across time within a fleet, then splitting by sex is adding parameters without adding information. The AT should examine whether this is the case. An alternative approach such as a hybrid-F method, which solves for sex-aggregated fleet-specific harvest rates iteratively given observed total catches and predicted vulnerable biomass, would achieve the same outcome with far fewer estimated parameters and would force the model to reconcile sex ratios through population dynamics and selectivity rather than through unconstrained F estimates. Under this framework, observed catch sex ratios could be fitted as a separate likelihood component with uncertainty that reflects the actual sampling coverage (e.g., Sentinel program sample sizes, buyer sampling availability, and mixed-bushel allocation rules), rather than being implicitly treated as known through tight fits to sex-specific catches. More broadly, the model estimates 2,639 parameters, and the ratio of estimated fishing mortality parameters to the information content of the data should be critically evaluated. The current model with its large number of parameters to be estimated may carry a risk of overfitting.

The fishery selectivity raises several questions. Selectivity was fixed based on minimum size regulations rather than estimated, which is a defensible simplification, but the iterative addition of constants to the inflection point parameter to improve fits introduces a degree of ad hoc tuning that is difficult to evaluate. Was this required due to the large size bins (1 cm) and the assumption of an even distribution of crabs across a length bin? For the surveys, the use of spline-based selectivity for the Maryland and VIMS trawl surveys produces flexible curves, but several aspects of the results require closer examination and justification. First, the report presents estimated selectivity splines for only three months for each survey (May–July for the MD Trawl, April–June for the VIMS Trawl), with selectivity fixed at 1 for all remaining months. The rationale for this choice “that censoring of recruits precludes estimation of the ascending limb in later months” is stated but not fully convincing, and selectivity for all months should be shown and the decision justified more transparently. Second, and more concerning, both surveys show strikingly similar dome-shaped selectivity patterns in the spring months, with reduced selectivity for large crabs. If these splines are capturing spatial availability driven by migration (mature females moving south toward the spawning grounds), then the patterns should be opposing, not parallel. Large crabs migrating from the upper Bay (Maryland) to the lower Bay (Virginia) should become less available to the MD Trawl and more available to the VIMS Trawl during the same months, not less available to both simultaneously. The fact that both surveys show the same declining selectivity for large crabs in spring suggests the splines may be compensating for a structural model-data conflict rather than representing genuine availability. This warrants investigation: are the splines absorbing misspecification in growth, natural mortality, or spatial population structure? Without a clear biological mechanism that explains why large crabs would be simultaneously unavailable to trawl surveys in both the upper and lower Bay, the estimated selectivity shapes should be treated with caution. More fundamentally, the model does not distinguish between encounter probability (does the gear contact the animal?) and retention (given contact, is the animal caught?). For the commercial fisheries, the assumption that recreational selectivity is identical to commercial selectivity for each jurisdiction is untested and likely incorrect given the different gears and fishing behaviours involved.

The model operates on a monthly time step throughout the year, but this level of temporal resolution may be unnecessarily complex during periods when little is happening biologically. During the winter months (December–March), there is no growth, no reproduction, and no commercial fishing other than the now-closed Virginia dredge fishery. Pooling these months into a single time step could simplify the model without sacrificing biological realism. The AT could explore a quarter time step for the model, which balances biological realism and model complexity.

Several structural features that could improve model performance were not explored. First, the model does not include process error. All variability is attributed to observation error, which means the model is forced to track every signal in the data rather than allowing for unmodelled stochasticity in population dynamics. Adding process error on abundance or recruitment would provide a more realistic characterisation of uncertainty and could improve the residual patterns observed in the early years of the time series. Second, the model's initial conditions assume quasi-equilibrium in 1994, but the large standardised residuals in the mid-1990s suggest this assumption may be inappropriate. Exploring alternative burn-in approaches or extending the model backward using available pre-1994 catch data could help resolve this. Third, the size distribution of recruits is assumed constant over the entire time series, yet changes in the timing of settlement, environmental conditions, or nursery habitat quality could shift both the mean size and variability of recruits entering the population. Allowing non-stationarity in recruitment size parameters would provide flexibility to accommodate such changes and could act as a proxy for temporal variation in both the size and timing of recruitment, given the simplification of modelling it as a February pulse. Fourth, the model does not include any explicit density-dependent natural mortality (cannibalism), despite cannibalism being well-documented in blue crab ecology and potentially important for understanding the relationship between recruitment and subsequent juvenile survival.

The model is spatially aggregated, treating the entire Chesapeake Bay as a single population unit despite substantial spatial heterogeneity in blue crab demographics. The WDS length-frequency data, when separated by state, reveal strikingly different population structures. Maryland stations show a population dominated by small crabs with a secondary mode (~130 mm) of medium-sized animals, while Virginia stations show a bimodal distribution with a much more prominent mode of large crabs (~145 mm), bigger than those in Maryland. These differences reflect the known spatial segregation of life stages, with juveniles concentrated in upper Bay nursery habitats and mature females migrating to the lower Bay for spawning. Fitting a single selectivity and catchability to the Bay-wide WDS index forces the model to average across what are essentially different demographic components. A more informative approach, as discussed under Tor 2, would be to treat spatial sub-regions of the WDS as separate fleets in the model, each with its own selectivity. This allows the model to fit the distinct size compositions observed in different parts of the Bay and better represent spatial population structure without requiring a fully spatial model.

The size-varying natural mortality applied to crabs ≤ 5 cm adds complexity in a size range where the data may be most limited due to the sampling limitations of the surveys (minimum depth of 1.5 m for WDS, gear selectivity for trawl surveys). The model could potentially be simplified by recruiting at a larger initial size (e.g., 5 cm, still age-0+), which would reduce the need for the size-dependent M scaling and place the model's dynamics in a size range better supported by the available data.

Finally, the data weighting approach deserves scrutiny. Effective sample sizes for composition data were determined using the number of sampling trips or fixed values, with a weighting term of 0.25 applied to non-WDS surveys chosen such that the model fit the WDS index well, and survey standardised residuals were manageable. This approach, while practical, lacks the iterative self-consistency of established methods such as Francis weighting, which re-weights composition data based on the fit of the model to the mean length or age. The assumed log-scale standard deviations for catch (0.1) and survey indices (0.25 for WDS, 0.4 for others) were fixed rather than estimated or iteratively tuned, and the report does not discuss whether these values appropriately balance the information content of the different data sources. Bias adjustment or ramping on recruitment deviations, which is standard practice in models like Stock Synthesis to avoid biased estimates in the early and late parts of the time series, does not appear to have been considered. Given the sensitivity of the model to growth parameters documented in the sensitivity analysis, and the large number of fixed assumptions in the model, a more systematic approach to data weighting and model diagnostics would strengthen confidence in the assessment results. The peeler fleet, for which catch can be reported directly in numbers of crabs, could be fitted in numbers rather than weight, avoiding the additional uncertainty introduced by weight conversion factors.

TOR 5: Characterize uncertainty in assessment estimates (mortality and abundance).

Panel Decision: Partially met.

The Panel evaluated whether the assessment identified and quantified uncertainty in key outputs, particularly for the estimates of abundance (males, immature females and females), fishing mortality F, and recruitment.

The Panel agreed that the AT identified major sources of uncertainty, including (1) observation uncertainty consisting of survey sampling errors and catch and effort reporting error; (2) process uncertainty such as interannual variability in recruitment, natural mortality, growth, maturation, and survey catchability; (3) model/structural uncertainty with assumptions about constant or simplified natural mortality, catchability (q), and (lack of) stock-recruit relationships; and (4) data gaps such as discard mortality, recreational harvest, and environmental drivers.

The AT quantified some of these uncertainties in the assessment, including uncertainties associated with survey data, commercial and recreational fisheries data, and life history parameters. The AT described the major uncertainty sources in an explicit and transparent way. However, the uncertainties in the assessment largely reflect sampling error only, and do not fully include catchability uncertainty, environmental variability, and process error. No time-varying life history parameters (e.g., M, recruitment and growth) and no environmental drivers were explicitly considered. No fully integrated uncertainty framework such as state-space or Bayesian treatment was explicitly considered, and there is a limited accounting for correlated errors across processes and possible interactions among uncertainty sources.

The AT conducted sensitivity analyses with respect to natural mortality assumptions, recruitment variability and some survey inputs, which helped bound uncertainty where direct estimation was difficult and provided insight into model robustness. However, the process of developing the base model and sensitivity analyses was not well structured and not well documented, making it difficult to have a full understanding of the uncertainty sources considered in the assessment and whether these uncertainties have been fully explored and evaluated.

The Panel considered the AT clearly and thoroughly identified uncertainty sources, had strong treatment of survey-based (sampling) uncertainty, and used sensitivity analyses to explore key assumptions. However, explicit quantification of process uncertainty such as natural mortality uncertainty and catchability uncertainty is limited.

The Panel recommends that the AT develop state-space or Bayesian frameworks to separate process versus observation errors and explicitly consider time-varying mortality uncertainty and integrate uncertainty across surveys, catch data, and environmental drivers. The Panel recommends that the AT document the protocols and process that led to the development of the base model and sensitivity analysis models.

TOR 6. Update the sex-specific catch survey models used in the 2011 benchmark stock assessment with relevant new data. Characterize major changes in assumptions between the 2011 assessment model and the 2023 model.

Panel Decision: Fully met.

The differences in catch accounting between the two assessments are substantial. The 2011 assessment did not include incidental mortality, abundance (number of crabs caught in the fishery) was calculated directly from catch weight, recreational harvest was a stationary proportion of total catch, and the removal time series started in 1968. The current assessment uses fleet and sex specific incidental mortality, abundance is calculated via a length structured model, recreational catch is sex and jurisdiction specific, and the removal time series starts in 1994 when reporting became mandatory.

Despite the differences in the way catch is dealt with in each assessment, the patterns of removals were highly correlated.

There were substantial differences in the way fishery independent survey data was incorporated as well. In the 2011 assessment, the WDS was included as an annual absolute index of abundance with catchability determined outside the assessment model. VIMS trawl data was included, but with a restricted spatial domain and design-based indices were used for most surveys (excepting the Maryland Trawl Survey). In the current assessment, the WDS provided an index of abundance. Relative catchability between WDS platforms were based on estimates from VAST using paired tows and areas of spatial overlap. Survey indices were monthly and model based (VAST).

Survey indices between the two assessments were similar in trend, but not strongly correlated in most cases.

Updating model runs from the 2011 assessment was difficult in part because updates to the ADMB optimizer introduced an instability in the estimation of the stock recruitment function. The AT built a STAN based version that converged reliably and used it for further model continuity comparisons.

Model continuity runs were: 1) updated 2011 assessment (SSCMSA) run using an old version of ADMB to get convergence. 2) STAN version of SSCMSA run through 2023. 3) Comparison between updated SSCMSA and STAN version of SSCMSA, run from 1994 – 2023 to match the current assessment time stanza and using model-based survey indices. 4) Comparison between updated SSCMSA and current model (SLAM).

The updated SSCMSA and the STAN version using the same time stanza produced similar parameter estimates in general, although the stock recruitment parameters (which caused convergence issues in modern ADMB) were somewhat different. Trends in recruitment, exploitation rate, and abundance were similar.

When model-based survey indices were used in the STAN version of SSCMSA, the male abundance and exploitation rate trends began to drift away from the updated SSCMSA version somewhat, although female trends and recruitment estimates remained similar.

Updated SSCMSA and SLAM exhibited similar trends in abundance but different scales, with the SLAM estimates of abundance being almost double those of SSCMSA.

TOR 7: Based on assessment model results recommend appropriate biological reference points for management. To the extent possible, evaluate the appropriateness *and utility of (1) Aggregate bay-wide reference points; (2) Sex specific reference points; and (3) Recruitment reference points.*

Panel Decision: Partially met.

The AT successfully derived BRPs from model outputs, including abundance-based, mortality-based thresholds/targets reference points, and recruitment-based reference points. The reference points are sex-specific. Their biological realism, management utility, and consistency with stock dynamics were well considered.

Because there is no assumed stock-recruitment relationship, F40% was used as the F threshold reference point, and 75% of F40% was used as the target F reference point. The median value of recruitment over the whole time series was used to calculate abundance-based target and threshold reference points from the relevant F reference points. The estimation was relatively simple and transparent. However, the appropriateness of the BRPs has not been quantitatively evaluated. The use of the median recruitment of the whole time series may not be suitable given large changes in the environment over the time. It may be more relevant and appropriate to use more recent recruitment in the estimation of abundance reference points.

The development of a recruitment reference point is novel, and provides an early indicator of future stock status. However, recruitment has extremely high interannual variability, and may be influenced by strong environmental forcing, which may make its interpretation and utility complex. The use of a 3-year moving average will help smooth any large annual fluctuations in the recruitment estimate.

Although the Panel agreed that the current biological reference points are well defined and justified, they are still ad hoc reference points and have not been quantitatively evaluated for their effectiveness in achieving management goals. The Panel recommends that the biological reference points be quantitatively evaluated in a simulation setting such as management strategy evaluation. The Panel encourages the AT to continue developing recruitment-informed indicators, not necessarily strict BRPs, but early warning indicators and environmentally-linked triggers. The Panel also encourages the AT to explore dynamic BRPs to incorporate possible environmental changes and regime shifts.

TOR 8: Evaluate stock status relative to recommended reference points.

Panel Decision: Fully met.

The AT provided status determinations for male and female adult blue crabs as well as recruits. These values are not current as the assessment time series ended in 2023. Based on the benchmark assessment and the outdated time series, neither adult males or adult females would be experiencing overfishing. Adult females would not be overfished. Adult males would be overfished. Recruits would be below the theoretical threshold value.

TOR 9: Identify relevant ecosystem and climate influences (such as habitat, environmental drivers, prey availability, and predation/cannibalism) on the population dynamics and fisheries and, to the extent possible, explore other analyses that support the assessment.

Panel Decision: Partially met.

The AT did excellent work on several aspects of this TOR. Predation by blue catfish, hypoxia, over-winter mortality (OWM), and the effects of temperature on growth were thoroughly examined. A parallel study is underway to examine oceanographic effects on recruitment, but that project is not completed and was not part of this review. Cannibalism and density dependent mortality were not examined in this assessment. The panel notes that this TOR is open ended and because potential ecosystem and climate influences are complex and numerous, it would be very difficult to fully meet over the time horizon allotted for development of this assessment.

The predation effect of blue catfish was examined by comparing survey CPUE for both blue catfish and blue crabs in CB and its tributaries. Blue crab abundance was negatively correlated with blue cat fish abundance. The AT estimated that blue crab abundance would be approximately 2.8% higher bay-wide, on average, in the absence of blue catfish.

Hypoxic effects were examined by comparing model-based estimates of the annual hypoxic volume to the blue crab survey indices from September ChesMMap and lagged WDS stations using GAM. The results from this analysis were inconclusive.

The AT examined dead crabs caught in the WDS in an attempt to estimate OWM rates. They used a Bayesian linear mixed model to estimate annual length and sex specific OWM rates. These rates were then used directly in the assessment model. In general, OWM rates are higher during colder winters.

The AT spent considerable time attempting to model growth as a function of temperature. Unfortunately, these efforts were not ultimately successful as temperature based growth did not adequately explain growth in the spring and late summer.

Cannibalism is an important source of mortality for blue crabs, but was not explicitly considered.

The drivers of low recruitment in recent years remain unexplained, although there is some possibility that the ongoing study of oceanographic effects may help explain it.

TOR 10: *Identify existing data sources and gaps, and, to the extent possible, characterize the uncertainty in the relevant sources of data.*

Panel Decision: Fully met.

The Panel agreed that the AT thoroughly described the existing data gaps and uncertainty associated with the assessment.

Blue crab growth is an important source of uncertainty in the assessment model. The model is length structured and highly sensitive to changes in growth. Growth depends on maturity which is likely to vary spatially and with size. There was not sufficient data to fully describe maturity, or to allow growth to vary over space and time.

Blue crab recruitment drivers remain unresolved. Recruitment has been low in recent years despite near target adult female abundance. Possible drivers of this pattern include predation, including cannibalism with attendant density dependent mortality implications, loss of habitat for juvenile crabs, as well as various oceanographic factors. The AT looked into some of these factors and the investigation into blue catfish predation was particularly well developed, but drivers of patterns in recruitment remain largely unknown.

Catch of blue crabs is highly uncertain. The characterization of catch is typically coarse, with sex and size composition poorly described. There is very little data of any kind on recreational catch, which is thought to be an important component of the fishery. Estimates of discards are likely to be lower bounds at best since they are based on proportions of dead crabs in pots and do not include additional mortality of discarded live crabs.

The model was not able to explain the differences in size composition between surveys with selectivity alone. There are likely spatial patterns in survey catch that remain unaccounted for by bay-wide estimates of recruitment and growth.

Like many assessments, natural mortality remains uncertain. Data from a tethering study provided the basis for natural mortality in this assessment. Other sources of data, including an additional tethering study and tagging data were either not yet available or difficult to separate from fishing mortality.

The most synoptic and complete survey time series, the WDS, provided some additional source of uncertainty. The recruitment size distribution was modified in order to accommodate the pulse recruitment assumption in the model and selectivity was assumed rather than estimated. Despite a series of depletion studies, the capture efficiency of the WDS dredge remains uncertain. Previous estimates were considerably higher than the catchability estimated in this assessment. Low estimates of q in this survey may be a result of the WDS not sampling shallow water habitat. There has also been some drift in the survey design that has not been thoroughly documented.

TOR 11: Report on the status of research recommendations from the most recent benchmark assessment. Identify and prioritize research recommendations for future work.

Panel Decision: Fully met.

The AT provided a thorough list of research recommendations. Prioritization of research recommendations has been and will continue to be, a common task of the CBSAC. Many members of the AT also serve on the CBSAC and prioritization is probably better undertaken there. The research recommendations from the AT presentation are reproduced below.

TOR 1

- Improve understanding of growth either through lab or field methods (preferably field methods that can track individuals over time)
    - Understand temperature and growth relationship throughout the entire year
    - Potential analysis: tag recapture study based on unique marking similar to finger prints.
- Better understand how many crabs are in shallow waters that are not surveyed by the WDS
- Improve estimates of M

    - Potential analysis: Combine catch data with WDS data and tagging data to estimate changes in mature female natural mortality
- Improve understanding of the size-specific and monthly patterns of female maturation and female growth at maturity and whether these processes vary spatially
- Improve understanding of the spatial patterns of recruitment and how they have changed over time
    - Potential analysis: develop a VAST model to combine MD Trawl, VIMS Trawl, and ChesMMAP to estimate spatial recruitment patterns over time

TOR 2

- Standardize biological data collection across surveys (e.g., record pre-pubertal females across all surveys)
- Evaluate alternative ways to get length compositions and indices from fixed site surveys like the MD Trawl survey, or expand fixed site surveys to include random stratified sampling locations
- Reconsider design and estimators used for the WDS. The WDS now has a substantial amount of information that can be used to evaluate modified designs
- Evaluate the potential consequences and need for the current spatial sub stratification in WDS
- Validate SERC methods for identifying pre-pubertal females and determine the time between when females begin to be identified as pre-pubertal females and when they molt
- Reconsider removing recruits from the surveys in the fall, and develop an assessment model that allows for recruitment to occur over several months instead of just one
- Use a VAST model to estimate differences in catchability among surveys
- Collect blue crab carapace width, sex, and maturity data in the seine survey in Maryland, and continue to collect it in the VIMS seine survey

TOR 3

- Increase port sampling of commercial catch for length and sex composition
- Better understand discards and discard mortality in the commercial fishery
- Better understand recreational harvest
- Conduct independent validation of catch reporting
    - Potential analysis: Check regulation infractions with law enforcement
- Review sex- and market-specific unit conversions used by jurisdictions for harvest data

TOR 4

- Attempt to model recruitment with a stock-recruit function which would allow estimation of MSY-based reference points
- Develop a spatial model that allows for spatial patterns of fishing mortality and growth
- Explore a state-space modeling framework to develop a model with explicit incorporation of process errors and environmental drives
- Explore sensitivity of the modeling results with respect to different monthly growth patterns
- Evaluate the impacts of recruitment timing assumed for the model on the impacts the assessment results

TOR 7

- Develop and test harvest control rules to inform selection and use of reference points
- Evaluate the effectiveness of current reference points in achieving the management goals in a simulation setting such as management strategy evaluation

TOR 9

- Improve understanding of overwinter mortality of blue crabs in Chesapeake Bay
- Estimate effects of temperature on growth; determine if effects of climate change are apparent
- Examine the effects of marine heat waves on survival during molting at all stages from zoea to adults
- Expand spatial analyses to estimate the effects of blue catfish on blue crabs
- Projections of future salinities to forecast blue catfish expansion
- Determine the effect of hypoxia on blue crab habitat use

Panel Recommendations

- Better documentation of model selection process and development of base case and sensitivity runs.
- Spatial aggregation/disaggregation of surveys based on similarities in size composition and demographics. Consider the “fleets as areas” approach using these survey segments.
- Applying standardization techniques (eg. VAST, GAMLSS) to survey length composition data.
- Explicit analysis of environmental covariates in index standardization.
- Formal sensitivity analysis of the effect of size truncation in small crabs in some indices.
- Initiate collection of data on female maturity data in more surveys.
- Continue development of growth model, including spatial (and if possible temporal) variation in growth rates.
- Continue monitoring recruitment timing and its dependencies on environmental covariates such as temperature.
- Accounting of discarded live crabs and discard mortality rate could be valuable.
- Better monitoring of the sex and length composition of the VA landings data is needed.
- Landings validation and reporting error rates and types of errors would be valuable.
- Effort data from the pot/trap fisheries would be helpful.
- Estimates of sex-, month- and fleet-specific F may be masking conflicts in the data. Alternative approaches such as a sex aggregated hybrid-F estimation (solving for fleet specific harvest rates iteratively given observed catch and fishable abundance), could achieve similar outcomes with fewer estimated parameters and would force the model to reconcile sex ratios through population dynamics and selectivity rather than through unconstrained F estimates.
- The SLAM model may be over-parameterized. Number of parameters in light of the information content of the data should be considered.
- Decisions regarding selectivity should be formalized. Ad-hoc parameter tuning (fishery) and some months/surveys being fixed at 1 were not fully justified. The apparent conflict between what is believed regarding the migration of large female crabs to the spawning grounds in spring and the apparently similar dome shaped selectivity patterns in VIMS and MD trawl surveys was not explained.
- Making recreational removals a direct proportion of commercial landings implies they share a selection curve, which is probably not correct. This assumption should be tested.
- State space models should be explored. Adding process error to abundance and/or recruitment could improve residual patterns. Explicit consideration of correlated errors across processes would be an improvement.

- Extending the model back before 1994 could resolve problems with the assumed quasi-equilibrium assumption for 1994.
- Adding density dependent mortality to the model (cannibalism) would increase biological plausibility.
- This model could be improved by adding spatial components.
- Increasing the size at recruitment could simplify the model and reduce uncertainty around the least well described (smallest) size classes.
- Aggregation of the winter months should be considered.
- Alternative data weighting approaches should be considered.
- A formal harvest control rule should be developed.
- A formal MSE based on the recommended reference points should be considered.
- The use of median recruitment in calculation of the reference points should be reconsidered pending the outcome of the environmental drivers of recruitment study.
- When the environmental effects on recruitment study is completed, assess the stock-recruitment-environmental relationship.
- Survey (WDS and trawl) estimates of age-0 crabs should be obtained and compared to each other and to the model recruitment estimates.
- The relationships between age-0 crab estimates and abundance of adult crabs and catches should be explored.

Appendix 1. Realized agenda

2026 Chesapeake Bay Blue Crab Stock Assessment Review

April 13 - 16, 2026

Purpose: This meeting will convene modeling experts, state jurisdictions, and independent reviewers to evaluate the draft 2026 Chesapeake Bay Blue Crab Stock Assessment.

Agenda

Day 1: April 13, 2026

I. **Welcome & Introductions** **12:00pm – 12:15pm**
Brief introductions of the Chesapeake Bay Blue Crab Stock Assessment Committee & the review panel

Requested Action: Non-decisional
Lead: Dan Hennen (NOAA)
Materials: N/A

II. **Presentation of the 2026 Blue Crab Stock Assessment** **12:15pm – 3:30pm**
Review of TORs 4, 5, 7 and 8

Requested Action: Non-decisional
Lead: Michael Wilberg (UMCES)
Materials: Draft 2026 Blue Crab Stock Assessment Report

III. **General / Open Question Period** **3:30pm - 4:45pm**
Opportunity for questions from the review panel

Requested Action: Non-decisional
Lead: Dan Hennen (NOAA)
Materials: Draft 2026 Blue Crab Stock Assessment Report

IV. **Public Comment** **4:45pm -5:00pm**
Opportunity for public comment & questions
Requested Action: Non-decisional
Lead: Dan Hennen (NOAA)
Materials: Draft 2026 Blue Crab Stock Assessment Report

V. **Adjourn** **5:00pm**

<u>Day 2: April 14, 2026</u>

I.	**CIE Term of Reference Review and Discussion** Review of TORs 2 & 3 Requested Action: Non-decisional Lead: Dan Hennen (NOAA) Materials: Draft 2026 Blue Crab Stock Assessment Report	**9:00am – 1:00pm**
II.	**Lunch**	**1:00pm – 2:00pm**
III.	**CIE Term of Reference Review and Discussion (continued)** Review of TORs 1, 6, 9 Requested Action: Non-decisional Lead: Dan Hennen (NOAA) Materials: Draft 2026 Blue Crab Stock Assessment Report	**2:00pm - 4:45pm**
IV.	**Public Comment** Opportunity for public comment & questions Requested Action: Non-decisional Lead: Dan Hennen (NOAA) Materials: Draft 2026 Blue Crab Stock Assessment Report	**4:45pm -5:00pm**
V.	**Adjourn**	**5:00pm**

Day 3: April 15, 2026

I. CIE Term of Reference Review & Discussion **9:00am - 10:00am**
Review of TORs 10 & 11
Requested Action: Non-decisional
Lead: Dan Hennen (NOAA)
Materials: Draft 2026 Blue Crab Stock Assessment Report

II. Review Session [closed-door] **10:00am - 12:00pm**
Opportunity for the review panel to discuss assessment methodologies and develop individual opinions. The review panel will initiate development of the panel summary report.
Requested Action: Non-decisional
Lead: Dan Hennen (NOAA)
Materials: Draft 2026 Blue Crab Stock Assessment Report

III. Lunch **12:00pm - 1:15pm**

IV. Review Session (continued) [closed-door] **1:15pm - 3:00pm**
Opportunity for the review panel to discuss assessment methodologies and develop individual opinions. The review panel will initiate development of the panel summary report.
Requested Action: Non-decisional
Lead: Dan Hennen (NOAA)
Materials: Draft 2026 Blue Crab Stock Assessment Report

V. Plenary Session **3:00pm - 4:15pm**
Review panel discloses decision on TORs.
Requested Action: Decisional
Lead: Dan Hennen (NOAA)
Materials: Draft 2026 Blue Crab Stock Assessment Report

VI. Public Comment **4:15pm -4:30pm**

VII. Adjourn **4:30pm**

Day 4: April 16, 2026

I. Review Session [closed-door] **9:00am - 12:00pm**
Opportunity for the review panel to continue to discuss assessment methodologies and develop individual opinions. The review panel will continue the development of

summary documents and present the summary report. Next steps and timeline for finalizing the summary report will be identified.

Requested Action: Non-decisional
Lead:Dan Hennen (NOAA)
Materials: Draft 2026 Blue Crab Stock Assessment Report

II. Lunch **12:00pm - 1:15pm**

III. Review Session (continued) [closed-door] **1:15pm - 4:15pm**

Opportunity for the review panel to continue to discuss assessment methodologies and develop individual opinions. The review panel will continue the development of summary documents and present the summary report. Next steps and timeline for finalizing the summary report will be identified.

Requested Action: Non-decisional
Lead: Dan Hennen (NOAA)
Materials: Draft 2026 Blue Crab Stock Assessment Report

IV. Public Comment **4:15pm -4:30pm**

Opportunity for public comment & questions
Requested Action: Non-decisional
Lead: Dan Hennen (NOAA)
Materials: Draft 2026 Blue Crab Stock Assessment Report

V. Adjourn **4:30pm**

Attendees

Adriana Celtruda (UMD)
Alexa Galvan (VMRC)
Alexei Sharov (MDNR)
Amanda Bevans (MSU PEARL)
Amy Schueller (NOAA)
Brooke Lowman (VMRC)
Bruce Vogt (NOAA)
Carrie Kennedy (MDNR)
Christina Garvey (CRC)

CJ Schick (SC DNR)
Dan Hennen (NOAA)
Danielle Kozkinski (MBA)
Dong Liang (UMCES CBL)
Gabrielle Salula (VIMS)
Gina Ralph (VIMS)
Graham Wagner (SC DNR)
Heather Hayden (MDNR)
Holly Funkhouser (SC DNR)
Ingrid Braun-Ricks (PRFC)
Jeff Brunson (SC DNR)
Madison Sholes (UMCES CBL)
Mandy Bromilow (MDNR)
Matt Ogburn (SERC)
Matthew Damiano (NC DMF)
Matthew Zink (NCDEQ)
Maya Drzewicki (UMCES CBL)
Michael Wilberg (UMCES CBL)
Nick Caputi (DPIRD/CIE)
Nick Staten (CRC)
Rob Aguilar (SERC)
Rob Latour (VIMS)
Rom Lipcius (VIMS)
Ron Owens (PRFC)
Sharon McBreen (PCT)
Simon de Lestang (DPIRD/CIE)
Tom Ihde (MSU PEARL)
Tom Miller (UMCES)
Trey Mace (MDNR)
Troy Tuckey (VIMS)
William Stoker (MDNR)
Yong Chen (Stony Brook University/CIE)