Chesapeake Bay Blue Crab Stock Assessment Review 2011

29 – 31 March, 2011, Baltimore, Maryland

On behalf of the Center for Independent Experts (CIE)

Dr Julian Addison
Le Mas du Coq
946 Chemin du Puits des Gavottes
84300 Cavaillon
France

julian@masducoq.com

6 May 2011

Executive Summary

The 2011 assessment for the Chesapeake Bay blue crab fishery uses a sex-specific catch multiple survey analysis (SSCMSA) model. This is a major improvement on the 2005 assessment primarily because of the integration of the reference point calculations within the model using per recruit analyses. Previously, the population model and the calculation of reference points were two separate processes which were not necessarily consistent. The two main differences from the 2005 model are the inclusion of a renewal function in the form of a stock–recruitment relationship, and the incorporation of separate components for each sex, which allows the generation of sex-specific reference points. This is particularly important considering recent management changes which have been geared to conservation of the female stock and which have consequently changed the sex-specific exploitation rates.

The assessment is a two stage (age-0 and age-1+ crabs) model, which incorporates three fishery-independent surveys including the extensive winter dredge survey, which is treated as an absolute index of abundance. Two main concerns with the formulation and robustness of the model were identified. Firstly, the stock-recruitment relationship is based on both males and females, and whilst it is an innovative approach, the way in which the relationship is formulated within the assessment model suggests that some recruitment will always occur even if male population size is zero. Secondly, the assessment model appeared to be extremely sensitive to treating the winter dredge survey as a relative abundance index, an approach conventionally used when incorporating survey data. Such an approach appeared to have significant implications for estimating stock status, reference points and stock status relative to reference points, and therefore it should be a priority to investigate the basis for treating the winter dredge survey as an absolute index given that the best model fit is treating the survey as a relative index of abundance.

The derivation of reference points considers the sex-specific formulation of the stock-recruitment relationship and whilst the use of US Federal and Regional Management Council regulations to define the limit and target reference points is supported, the assessment makes a value judgement about the overfished limit and target reference point rather than give a range of potential values, and the value is not as precautionary as it could be. This is important because the model has been developed as part of a benchmark assessment of the stock and there will not be a new assessment each year. Instead in future years, empirical estimates of exploitation rate and abundance will be calculated from observed catch levels and abundance based on the winter dredge survey and these will be used to evaluate the status of the stock in relation to a control rule based on exploitation rate and abundance reference points. It is important therefore that any inconsistencies in the model structure are ironed out before the control rule is applied to future empirical estimates of exploitation rate and abundance.

A number of other areas were highlighted where the assessment could be improved. The assessment would benefit from the incorporation into the model of additional fishery-independent surveys, the inclusion of the Potomac Fisheries Commission time series of fisheries data, a consistent standardised catch rate dataset that can be used in the assessment, an updated assessment of annual levels and variation in recreational catch, and some quantification of incidental mortality in the fishery. There is also potential scope for refining and improving the model through the incorporation of size classes to more fully represent blue crab biology, time steps of less than one year, i.e. a quarterly or even monthly model, and possibly incorporating the information on soft/peeler crabs alongside the hard crab fishery as separate entities, and the incorporation of spatial structure in to the model.

Background

Blue crab is the most important commercial fishery in Chesapeake Bay. The first Baywide stock assessment (Rugulo *et al*., 1997) was conducted using a length-based approach to estimate exploitation, and an unweighted average of the four principal fishery-independent surveys to determine abundance. Consequently biological reference points were crude. In 2001, the Bi-State Blue Crab Advisory Committee (BBCAC) developed a new management framework that proposed threshold reference points based on maintaining 10% of the virgin spawning potential and on the lowest observed abundance, and a target exploitation rate that would lead to an effective doubling of the spawning stock present in 2001. In 2005 a second Baywide benchmark assessment for blue crab was conducted (Miller *et al*., 2005) and recommended adopting the exploitation fraction, defined as the proportion of crabs available at the beginning of the season that are subsequently harvested, in place of less intuitive measures (F) used in previous assessments. Estimates of exploitation fractions were calculated based on the Baywide winter dredge survey (WDS) and within a modified catch-survey analysis that permitted the use of multiple surveys. The approach used in the 2005 assessment was reviewed by a panel of international scientists with expertise in crustacean fisheries who found that it was a substantial improvement over previous assessments (Haddon, 2005) but also identified issues to be addressed in future assessments.

Since the 2005 assessment, the three management jurisdictions have implemented a range of regulatory changes aimed at attaining the target exploitation rate of 46% of the available stock. In 2009 the NOAA Chesapeake Bay Office initiated and supported development of another benchmark assessment and that assessment (Miller *et al*., 2011) is the basis of this CIE Peer Review.

Description of Review Activities

The CIE review for the Chesapeake Bay blue crab assessment took place in the Sheraton Hotel, Baltimore, Maryland on 29-31 March 2011. In addition on 28 March there was a special session at the National Shellfisheries Association Annual Meeting devoted to the results from the Blue Crab Advanced Research Consortium which provided interesting background to the fishery to the review panel but which was not formally part of the review process. Prior to the review, the NOAA Chesapeake Bay Office in conjunction with the University of Maryland, Chesapeake Biological Laboratory provided all the documents on a website for the Review (Appendix 1). Following a request for the code in a form that could be run in ADMB and R, these were kindly provided to the review panel via the website.

The objective of the CIE Chesapeake Bay blue crab review was to review and discuss input data, assumptions and parameters to the model, the assessment itself and the biological reference points for the population.

The Terms of Reference were:

a) Critically assess and where necessary revise the life history and vital rates of blue crab in the Chesapeake Bay that are relevant to an assessment of the stock.
b) Evaluate and recommend biological reference points for the Chesapeake Bay blue crab population. The potential for implementing sex-specific reference points should be evaluated.

c) Describe and quantify patterns in fishery-independent surveys. Analyses should include an evaluation of the impacts of environmental and abiotic factors on survey catches, to maximize the information content of resultant survey time series.
d) Describe and quantify patterns in catch, effort and survey-based estimates of exploitation by sector and region, including analyses that examine the impacts of reporting changes and trends in CPUE.
e) Develop and implement assessment models for the Chesapeake blue crab fisheries. In particular, models that permit estimates of the trends and status of the crab population and fisheries on a sex-specific basis should be evaluated.
f) Examine density-dependent exploitation patterns derived from survey-based and model-based approaches.
g) Characterize scientific uncertainty with respect to assessment inputs and stock status.
h) Evaluate stock status with respect to reference points.

On the first day of the review meeting (29 March), Dr Tom Miller, the lead investigator on the assessment, gave a presentation covering all aspects of the above Terms of Reference (ToR), and there was a wide-ranging discussion between the assessment team, the review panel, other interested scientists and representatives from the relevant management jurisdictions. On the second day (30 March), the review panel discussed the research with the assessment team against each ToR. Throughout the review meeting, the whole assessment team was extremely helpful and flexible in critically discussing all aspects of the assessment and the ToRs of the review, which made the task of the review team immeasurably easier and more enjoyable. Upon request, the assessment team provided the following additional documentation, information and tests:

1. Correction of the lagged plots;
2. Extra residual and predicted versus observed plots;
3. An additional preliminary model run assuming that the winter dredge survey is a relative index of abundance.

The review panel met in private on 31 March to evaluate progress in relation to all the ToRs and in particular to evaluate the assessment and the consequent derivation of biological reference points for the population. The panel also met briefly on 1 April to review some additional material provided by the assessment team and to complete a summary report.

Summary of findings

The findings of the review are structured in line with the eight Terms of Reference (ToRs) provided to the review panel, although many aspects of the assessment discussed at the review related to more than one ToR, and this is reflected in the summary of findings below.

ToR a Critically assess and where necessary revise the life history and vital rates of blue crab in the Chesapeake Bay that are relevant to an assessment of the stock.

Over the last decade, significant progress has been made in understanding the life history and vital rates of blue crab in Chesapeake Bay, which has improved the reliability of parameter estimates for input to assessment models. The initial question that is relevant to an assessment of the stock in Chesapeake Bay is whether the Chesapeake Bay population can be considered as a single stock. Genetics studies suggest that it is a reasonable to assume that the

Chesapeake Bay population is a single stock for fisheries management purposes, and whilst larvae studies suggest some limited exchange with other areas, they generally confirm this exchange is not expected to significantly affect population dynamics of blue crab in Chesapeake Bay. In comparison with most fish "stocks", I agree with the conclusion that the Chesapeake Bay stock can reasonably be considered as a separate stock.

In terms of the assessment model, the key biological parameters are the growth rates, reproductive parameters and natural mortality rates of the adults.

Growth rates. This seems to be a key area for the assessment in relation to the method used to allocate crabs in the surveys to the two stages (age-0 and age-1+ crabs) for the population model. In addition, these two stages form the basis of estimates of recruitment, exploitation rate and spawning stock reference points with age-0+ crabs providing exploitation rate reference points, whereas age-1+ crabs provide spawning stock /abundance reference points. An important question is how much in-year dynamics might influence the recruitment of crabs to the fishery, and whilst the assessment team were confident that empirical observations from the surveys show that the cut-off points for age-0 and age-1 crabs are very clear and do not vary significantly from year to year, I would recommend that annual variations are investigated to ensure that assumptions underlying the population model are correct.

Growth information comes from the use of estimates of age based on accumulation of lipofuscin in neural tissue pioneered for the blue crab by Ju *et al*. (1999, 2001) but significant questions concern the use of this technique. Firstly the technique has been criticised as being inappropriate for estimating the age of long-lived animals and a study on freshwater crayfish (Fonseca *et al*., 2003) showed that lipofuscin concentration could under certain circumstances decline with time (age of animal). In the case of the blue crab this should not be a problem because the species is relatively short-lived. In particular, however, there has been considerable disagreement about the methodology of measuring lipofuscin concentration by employing spectrofluorimetry of crude tissue extracts (Harvey *et al*., 2008; Sheehy, 2008). However the recent study by Puckett *et al*. (2008) using known-age blue crabs from a hatchery suggests that greater confidence can now be attached to the estimated growth rates and subsequent partial recruitment curves for blue crabs. Nevertheless there remains some uncertainty around these parameter estimates, and in particular I would recommend some sensitivity analysis of the assessment model outputs to the partial recruitment curve generated from the Puckett *et al*. study.

Reproduction. Little is known about age and size at maturity in blue crabs, but mating and spawning periods are well understood. Female crabs produce up to 8 clutches per breeding season, but viability is lower in later clutches, so most recruitment comes from the initial clutches. Larger female crabs produce larger, but fewer clutches, but no information was presented on size and viability of eggs in relation to the size of female crab. The larvae cycle is well understood and empirical observations suggest that the sex ratio at recruitment is 52% female.

Whilst there is a wealth of knowledge about the reproductive biology of blue crabs, and more is becoming available through a concerted research effort in conjunction with a hatchery and stock enhancement programme in recent years, reproductive parameters are handled relatively simply within the population model. In the population model, spawning stock estimates seem to be based solely on numbers of age-1+ female crabs, and there would appear to be scope to

make this spawning stock calculation rather more sophisticated by including fecundity-at-size and hence producing more accurate estimates of egg production. At present my main criticism of the way in which the model is formulated is that it does not appear to adequately accommodate the assumed terminal moult in blue crab females. It is not physiological because the Y-organ does not degenerate, but appears to be functional because females don't appear to moult after being inseminated. At present the population model assumes that all female crabs of age-1+ are identical, when in fact with a terminal moult, females will mate once, not moult again and hence are unlikely to contribute further to the spawning stock/egg production. On that basis, the assumption that spawning stock is proportional to the number of age-1+ females is incorrect because the larger, older females will not be contributing to the spawning stock. I would recommend that the population model is modified to more accurately assess spawning stock.

Whilst there is potential for sperm limitation in blue crab, recent monitoring has shown that all mature females appear to have sufficient sperm. However, some recent unpublished PhD work has suggested lower clutch sizes in females, and so there is clearly a need to investigate further the potential for sperm limitation in blue crabs, particularly in the light of recent new management measures to protect the female stock and which are therefore causing an increase in the male exploitation rates.

Natural mortality rate estimates

The only feature of the adult life cycle that is incorporated into the assessment is an estimate of natural mortality rate (M). Major effort has been put into estimating M since the first assessment in 1997 when M was assumed to be 0.375 (Rugulo *et al*., 1997). The previous stock assessment (Miller *et al*., 2005) used a range of values of M from 0.6 to 1.2, but also included the previous estimate of 0.375.

Since the last assessment, Lipcius and Smith (2011) have analysed returns from a tag-recapture study of mature female crabs using a Brownie-type model (Brownie *et al*., 1985) to estimate average annual survival rates of 0.15 varying from 0.09 to 0.28. Estimates of maximum age were then used with Hoenig's (1983) empirical relationship to give estimates of M ranging from 0.65 to 0.79. (A similar estimate of M could be calculated for the maximum age of crabs identified through the lipofuscin-based ageing approach, but this was not possible because the lipofuscin studies were undertaken at a time when only small, young crabs were available for sampling.) It would be instructive to investigate how the various components of mortality contribute to this estimate of 15% survival. As an initial step, use of the Brownie model with a different parameterisation that includes M and F, or M and catchability and effort would be informative. This would be a better method of calculating M than the method of using the maximum age calculated from the tagging experiment.

In comparison with most fisheries, the assessment team have collated a huge range of information on estimating the natural mortality, which provides added confidence that they are using a plausible range of values for M. The population model does however assume that M is stage and sex-independent and constant. Whilst there is no evidence that blue crabs are susceptible to very high incidences of mortality due to disease such as Hematodinium infection, a paper presented at the special session of the Blue Crab Advanced Research Consortium on 28 March (Kramer *et al*., 2011) showed that M can vary annually, seasonally and with crab size. It is recommended that an analysis is carried out to assess the sensitivity of the model's output to variable M.

In conclusion, major recent advances have been made on understanding the life cycle of the blue crab and its variation through the Blue Crab Advanced Research Consortium and the hatchery and stock enhancement programme, yet the population model used in the assessment remains relatively simple. For example, there is significant spatial structure of the stock within the whole bay/estuary, both east-west and north-south, in terms of habitat differences, mortality rates, prey and predator abundance, rates of cannibalism etc., and also in terms of movement of crabs around the bay including migratory corridors of mature females. Yet none of that spatial variability is incorporated in the model. I recommend therefore that a spatial component be added to the model or at the very least some spatial analysis of the survey data is undertaken. As with most hatchery and stock enhancement programmes, it remains to be seen whether enhancement could be a viable alternative to conventional fisheries management measures, or at the very least whether enhancement programmes could form part of an integrated management scheme. However, there is no doubt that the hatchery and enhancement programme is making major advances in the understanding of blue crab biology and ecology.

ToR b Evaluate and recommend biological reference points for the Chesapeake Bay blue crab population. The potential for implementing sex-specific reference points should be evaluated.

Reference points for the Chesapeake Bay blue crab fishery have evolved significantly since the first baywide assessment in 1997, and until now reference points have been estimated independently of the analyses used to assess exploitation rates and population abundance. The 2005 benchmark assessment derived reference points using the exploitation fraction (U) defined as the proportion of crabs available at the beginning of the season that are subsequently harvested. The overfishing definition equivalent to maintaining 10% of the virgin spawning stock was estimated as $U_{threshold}$ = 53% of all available crabs, and target exploitation rate (equivalent to maintaining 20% of the virgin spawning stock) was estimated as U_{target} = 46%. These reference points were used for management from 2006 onwards, but recently fishery managers, who were increasingly reliant upon abundance estimates from the winter dredge survey, became uncomfortable about the emphasis on exploitation rates alone, and so they adopted additional reference points of an abundance threshold of 86 million age-1+ crabs and an interim target abundance of 200 million.

However there has been concern that the reference points based on exploitation rate and abundance may not be consistent, so a key element of the latest (2011) assessment is the development of a fully integrated analysis that estimates both reference points and stock status. This is a major advance on the approach taken in previous benchmark assessments and is a real strength of the new assessment model. In addition the model permits the derivation of sex-specific exploitation rates and the authors recommend that all exploitation reference points should be based on age-0+ females, as some of the age-0 crabs will recruit to the fishery during the season. Secondly, all abundance-based reference points should be based on age-1+ females as these represent spawning stock and can be accurately measured during the WDS. This approach is fully supported.

The following reference points were defined from the 2011 assessment -

(1) Overfishing limit is defined as the exploitation rate of age-0+ female crabs that corresponds with MSY

(2) As blue crab is considered a data poor species (because there is no uncertainty attached to this estimate) and in line with regional Fishery Management Councils' regulations, target exploitation rate is set at 0.75 x U_{msy}.
(3) Overfished abundance threshold should be 0.5 x N_{msy}
(4) Target abundance reference point should be the equilibrium abundance when target exploitation rate is achieved, i.e. $N_{0.75 \times msy}$

I would fully support the approach taken in the assessment. The derivation of the reference point calculations considers the sex-specific nature of the stock-recruitment relationship and is implemented correctly in the model. The exploitation and abundance reference points are meaningful indices which are understood readily by all stakeholders. In particular, the exploitation fraction (U) is a much simpler concept to understand than reference points based on fishing mortality (F). This is particularly important if management action is required when thresholds are exceeded or if targets are reached and agreement is needed on how to achieve the management objectives. I support the use of the US Federal definitions for the limit reference points and the use of the New England and Mid-Atlantic Fishery Management Councils' system to define target reference points. These reference points may seem a little arbitrary, but by their very nature, most reference points have a degree of subjectivity, and the authors have sensibly chosen to define their reference points along the lines of widely-agreed approaches.

I shall leave discussion of the actual values for these reference points to a later section as they sit more easily alongside the results of the assessment under ToR e. I do have some concerns about the way in which the threshold and target abundance reference points have been estimated as the assessment team has made a choice that is normally the role of fisheries managers or policy makers, but again I will leave this discussion to ToR e.

Finally, I note that all reference points are now framed in terms of abundance of females, but with new management measures restricting the exploitation of females, recent increases in male exploitation rate and the increased potential for sperm limitation, it would be worth considering whether some male-based reference points might be appropriate, for example, in relation to abundance of age-1+ male crabs.

ToR c Describe and quantify patterns in fishery-independent surveys. Analyses should include evaluation of the impacts of environmental and abiotic factors on survey catches, to maximize the information content of resultant survey time-series.

The blue crab assessment benefits from having a series of long-running fishery-independent surveys. Three of these surveys, the VIMS trawl survey of the southern end of the Bay, the Maryland trawl survey of the north-eastern area of the Bay and the Baywide winter dredge survey (WDS), are used in the assessment, and other shorter time series survey data are available but were not used in the assessment. In the assessment, the abundance of age 1+ crabs in the WDS was assumed to be an absolute index of abundance, and other stages in the WDS and the other two surveys were considered as relative abundance indices.

In response to the peer review of the 2005 assessment, the authors adopted a delta generalised linear modelling approach (GLM) to obtain standardised indices of abundance for the three surveys. For all three surveys, the analysis showed that the best fitting models (lowest AIC) included survey design and environmental factors. For the WDS the best fitting models for

males, females and crabs of both sexes included all design and environmental factors. The use therefore of standardised abundance indices is a significant improvement on raw abundance data from the surveys. Standardisation of survey indices using the delta generalised linear model (deltaGLM) method is appropriate. However it is recommended that interaction terms in the second stage of the delta GLM be considered, especially strata and year, temperature and year for the dredge survey. However even after standardising the data, the correlations between abundance of age-0 crabs in year t with age-1+ crabs in year t+1 is relatively poor, suggesting that the allocation of crabs to age classes may not be as good as it might be due to inter-annual variation in growth rates.

It is clear that the fishery-independent surveys are a vital component of the assessment of the Chesapeake Bay blue crab stock. In particular, the WDS is a very important survey because it is an extensive baywide, stratified random survey, it takes place when the temperature is less than 10 degrees C and so crabs are buried in the substrate at these temperatures, year and vessel specific catchability are calculated for each year, and return visits are made to high density stations to estimate winter mortality. I would therefore strongly recommend continuation of the WDS as an essential component of the assessment and the evaluation of stock status relative to the reference points.

The other two fishery-independent surveys are of less importance, but I believe that more could be made of these survey data. For example, the Virginia (VIMS) trawl survey was not standardised and only 3 of the 7 strata were used in the trawl survey Index. The Virginia trawl survey should therefore be standardised and there are two options for making best use of the available data. The four strata that are not currently used could either be incorporated into a single index for the whole geographical area covered by the VIMS survey, or used as a second independent time series. The latter may be preferable because it would place greater emphasis on the winter component of the survey portion and potentially provide information on spawning abundance). Secondly, it should be possible to look at the WDS survey results in Virginia and Maryland separately and see how they correlate with trawl survey indices in the two states. There are also other survey time series, such as the Chesapeake Bay Multispecies Monitoring and assessment Program, which should be examined in more detail and potentially included in the assessment, and I would particularly recommend investigation of an earlier life history phase recruitment index, e.g. larvae or megalopae abundance.

ToR d Describe and quantify patterns in catch and effort by sector and region, including analyses that examine the impacts of reporting changes and trends in CPUE.

The blue crab fisheries in Chesapeake Bay are managed by three separate jurisdictions, Virginia Marine Resources Commission, Maryland Department of Natural Resources and the Potomac River Fisheries Commission and each management jurisdiction has historically evolved slightly different methods for collection of catch and fishing effort data, as well as having varying management legislation. Whilst pots, trot lines and dredges have been used consistently in the past, potting is the principal method of fishing now. Whereas biological and survey data for the fishery are excellent, there are significant doubts about the quality of landings and effort data from all areas.

One of the key problems in obtaining an accurate time series of catch and effort data has been changes in reporting procedures in both Virginia and Maryland over recent years. Following the study of Fogarty and Miller (2004), the current assessment reconstructed commercial

landings for Virginia and Maryland using time series analysis. Results show an impact of the 1993 reporting change in Virginia, suggesting that landings were under-reported by a significant amount prior to 1993, and this time series of adjusted landings suggests a significant recent decline in landings. There were also changes in reporting procedure in Maryland in 1981 and 1993 but analysis suggested that only the 1981 intervention was found to be significant. Again re-constructed time series suggests major under-reporting prior to 1981. Time series analysis of catch and effort trends is the standard accepted way of analysis to examine any effects of reporting changes. However the analysis of adjusting the catch to past reporting changes is very influential in the catch series used in the assessment and has a significant impact on the evaluation of the status of the stock. As it is possible that other concurrent events may have led to an increase in landings at the time of the change in reporting procedures, it is important that the scale of these changes be independently verified, even if indirectly and/or anecdotally.

The Potomac Fisheries Commission has the best time series of fisheries data because they have not been compromised by any changes in reporting procedures. However these data have not been analysed as part of the assessment because they are still on old style format discs. I strongly recommend that the assessment team gains access to, and analyses, these data and compares trends across other two states' surveys and the WDS.

There are a number of other areas where I would recommend further detailed investigation in order to provide a better understanding of removals from the fishery and potential areas for incidental mortality due to the fishing process.

First and foremost, I would recommend that accurate and complete catch and effort data are collected from the fishery so that a consistent standardised catch rate dataset can be used in the assessment. Such an index needs only to be representative, and it would be worthwhile investigating whether individual fishers had kept personal diaries of daily records of catch and effort. In addition, it would worthwhile reviewing the collection of size and sex data as sex-specific catches often have to be extrapolated across areas.

The assessment assumes that an additional 8% of commercial landings are due to the recreational sector, but recent studies by Aguilar *et al*. (2011) and Roberts *et al*. (2011) suggest that recreational harvest rates vary across sub-estuaries, and in some areas may be as much as 50%. I would recommend therefore that a baywide survey of recreational catches be undertaken immediately and at regular intervals to quantify the overall landings from the recreational sector. Similarly, it would be helpful to gain some quantitative assessment of under-reporting rates.

Finally, there are a number of aspects of the blue crab fishery that highlight the potential for large incidental mortality. The summer dredge fishery (currently closed) undoubtedly caused significant mortality to crabs, and there may also be high mortality rates during handling in the summer pot fishery. Some quantification both historically and in the current fishery could be included in the assessment calculations. The soft and peeler crab catch does not record catches at the point of removal from the water, but is instead based on landings of animals that have *survived* being kept in a shedding tank, and so the animals that died during this process have not been considered as a component of the catch. This requires either a) that the catch is recorded or b) the length of time they are kept in the tank is recorded and consequently the survival rate of these crabs is estimated. It would also be instructive to

obtain quantitative estimates of discard mortality rates and any potential "ghost fishing" due to lost pots.

ToR e Develop and implement assessment models for the Chesapeake blue crab fisheries. In particular, models that permit estimates of the trends and status of the crab population and fisheries on a sex-specific basis should be evaluated.

The 2011 assessment for the Chesapeake Bay blue crab fishery uses a sex-specific catch multiple survey analysis (SSCMSA) model. I consider that this is a major improvement on the 2005 assessment which used a catch multiple survey analysis (CMSA) model. The main differences from the 2005 model are the inclusion of a renewal function which allows recruits in year t+1 to be generated from abundance of adults in year t, so management reference points can now be generated in relation to both spawner per recruit and MSY. Previously, the population model and the calculation of reference points were two separate processes which were not necessarily consistent. Secondly, the model incorporates separate components for each sex, which allows the generation of sex-specific reference points and which is particularly important considering recent management changes which have been geared to conservation of the female stock and which have consequently changed the sex-specific exploitation rates.

The model uses standardized time series of fishery independent abundance and is fit to time series of total and sex-specific catches. In general, the model performs relatively well. The model replicated time series of total catch and the time series of the winter dredge survey (WDS), and the Virginia and Maryland trawl surveys. The model assumes that the abundance of age-1+ crabs in the WDS is an absolute index of abundance but catchability is estimated for other surveys. The best model fit estimated catchability of 0.4 for age-0 crabs in the WDS. However the model does not provide a particularly good fit to some of the data. The main areas of "lack of fit" are an inability to capture the high recruitment levels in many years, and secondly the model consistently underestimates the female population size and overestimates the male population size irrespective of input parameters. I would recommend detailed investigation into why the model provides such a poor fit to the sex-specific population sizes.

The model structure is considered to be appropriate (although see reservations below) but is not very well documented – both in terms of parameters, but also in terms of how it deals with life history events. I recommend that a conceptual model is presented before the mathematical components to allow readers to evaluate the assumptions of the model in relation to spawning, recruitment timescales etc.

The most important new aspect of the model is the introduction of a stock-recruitment relationship based on both males and females (equation 6.5). Whilst this is an innovative approach, this equation seems to suggest that some recruitment will always occur even if male abundance is zero! At current stock levels, this two sex SRR is acceptable, but we note that at extreme stock levels, the model does not perform in a realistic manner, and this should be fully documented in the assessment.

In addition there are a number of minor changes required to the formulation of the model identified by the review panel such as defining the initial conditions of the model,

decrementing M by time and F by k (kappa) in equations 6.7 and 6.11, incorporating a bias correction and incorporating a compensatory function into the stock-recruitment relationship.

Determination of reference points

The derivation of the reference points estimated from the assessment model have been discussed previously under ToR b, so in this section I will confine myself to discussion of the generation of the specific values below estimated for each of the agreed reference points.

- The overfishing limit is the exploitation rate of age-0+ female crabs that corresponds with MSY, U_{msy} = 0.34 based on a catchability of age-0 crabs of 0.4, and partial recruitment to the fishery of age-0 crabs of 0.6. This estimate of U_{msy} is relatively invariant to the ratio of the sex-specific exploitation rates.
- As the blue crab is considered to be a data-poor species and in line with regional management council regulations, the target exploitation rate is estimated at 0.75 x U_{msy} = 0.255
- The overfished abundance limit/threshold is calculated as 0.5 x N_{msy} =70 million age-1+ females. This figure of 70 million is taken from the lower end of the range of 68-82 million females estimated for varying sex-specific exploitation rates.
- Target abundance is defined as the equilibrium abundance when the target exploitation rate is achieved, i.e. $N_{0.75 \times msy}$ = 215 million age-1+ females. Again this figure of 215 million is taken from the lower end of the range of 210-340 million dependent on the assumptions about sex-specific exploitation patterns.

As stated under ToR b, I believe that the assessment team has taken a very sensible approach to defining reference points and this approach should be supported. For the abundance limit reference point, Regional Management Council guidance is to set this at 0.5 x N_{msy}. This is appropriate, but where I question the assessment team's approach is the choice of this limit reference point based on the range of sex-specific exploitation patterns assumed within the model. Figure 6.12 shows that N_{msy} occurs at a range of values dependent on the assumption underlying the sex-specific exploitation rates. For the observed range of relative exploitation rates, 0.5 x N_{msy} falls between approximately 68 and 82 million age-1+ crabs. However instead of giving this range of values from which fishery managers could choose their reference point, a more appropriate approach in my opinion, the assessment team have themselves made a decision to accept a value at the lower end of this estimated range at 70 million age-1+ crabs. Whilst it is open to discussion as to whether this is the assessment team's role, my main criticism is that this value of 70 million age-1+ crabs is in fact at the lower end of the predicted range, and therefore cannot be considered precautionary. A fully precautionary value for the overfished abundance threshold would be something around 82 million age-1+ crabs because this would be the estimated value for the highest sex-specific exploitation rates observed previously (a ratio of F_{male} to F_{female} of 1.6). In defining the target abundance, the authors have again chosen a specific value for $N_{0.75 \times msy}$ instead of providing managers with a range of options dependent on assumptions about the ratio of sex-specific exploitation rates.

Alternative stock assessment models

In addition to the SSCMSA model the assessment team have also provided the results of fitting two alternative assessment models. Production models fitted to time series of abundance and effort data in Working Paper 3 (Miller 2011a) suggest similar yields to the SSCMSA model, but abundance reference points were significantly greater than those from SSCMSA. Re-running the 2005 CMSA model (WP3, Miller 2011b)) did capture the trends in abundance of age-1+ females, but predicted much higher values of the exploitation rate (U). The assessment team should be congratulated on providing alternative assessment models, but the review panel agreed that differences in model structure in that female spawning stock size is not calculable from the production and CMSA models, comparing these with the new reference points is not possible, and concluded that the SSCMSA model was the best of the three models.

Finally, I would recommend that in the future consideration is given to incorporating more biological realism and structure to the model, particularly as so much more information is emerging from the major research programme carried out over the last 10 years. Potential options include incorporating shorter time steps such as quarterly or even monthly time steps to more explicitly incorporate timing of life history events, making the model size-based for better incorporation of fecundity–at-size and providing a better approximation of the reproductive status of females that undergo terminal moult, incorporating a spatial component to the model because of the widely-observed spatial differences, such as the very different nature to the fisheries in Maryland (primarily males) and Virginia (primarily females), and possibly incorporating the soft/peeler crab fishery alongside the hard crab fishery as separate entities.

I note that this is still very much a single species model and ultimately ecosystem-based models should be developed for the Chesapeake Bay blue crab population. However I would not recommend such a development as a priority at the current time.

ToR f Examine density-dependent exploitation patterns derived from survey-based and model-based approaches.

The assessment notes that temporal patterns of abundance (N) and exploitation rate (U= catch (C)/abundance) over time are mirror images of each other suggesting depensatory exploitation. The possibility of depensation was examined in detail in a working paper by Lipcius (2011). This elegant analysis calculated U from the annual estimates of baywide abundance (N) and the landings (C) from the three published Fisheries Commission statistics. His analysis showed that a depensatory relationship is likely with higher values of U at low population abundance, and so fisheries managers may need to consider additional management measures at low abundance.

Other density-dependent processes have been addressed in detail as part of the Blue Crab Advanced Research Program. For example, Johnson *et al*. (2011) observed density dependent movement of larvae, and also density dependent mortality attributed to cannibalism. Such density-dependent processes should be incorporated into the population model.

ToR g Characterize scientific uncertainty with respect to assessment inputs and stock status.

The assessment document provides a minimum of sensitivity analysis of the output in relation to the input parameters M, partial recruitment and sex ratio. I recommend therefore that a more systematic sensitivity analysis is undertaken to understand how robust the model's output is to these and other input parameters. In addition, new information is emerging about variation in some of these parameters. For example, Kramer *et al.*'s (2011) study showed annual, seasonal and size–related variations in natural mortality which should be incorporated in the model.

More importantly, however, I would like to see additional runs to investigate the sensitivity of the model to various aspects of the model's configuration and some of the key assumptions underlying the model. These might include running the model with and without the modified time series of catches, relaxing the assumption that abundance of age-1+ crabs are an absolute estimate of abundance, runs where observation error and process error are not estimated simultaneously, and alternative configurations of the sex-specific stock recruitment relationship.

One of the key assumptions of the model is that the abundance of age-1+ crabs from the WDS is an absolute estimate of abundance, whereas the model estimates catchability of age-0 crabs from this survey. In most stock assessment models, it is conventional to assume that all survey indices are relative rather than absolute indices of abundance, and so the panel requested the assessment team to carry out additional runs where catchability of the age-1+ crabs was estimated from the model rather than assumed to be 1.0. Preliminary runs of the model showed a much better fit to the sex-specific abundances, but still failed to predict the high levels of recruitment. Estimates of catchability appeared to be very different for this model than the baseline run and the catchability estimates may change the interpretation of the reference points and trajectory of the stock. The model is therefore extremely sensitive to treating the WDS as a relative index of abundance in terms of absolute stock status, reference points and stock status relative to reference points. As the assessment model is not robust to this assumption, the panel recommended that a sensitivity test is run that uses the raw winter dredge survey indices and that the derivation of the catchability of the winter dredge survey index of abundance should be investigated and possible sources of this difference (both assessment and survey) should be investigated.

I recommend therefore that a full investigation of the behaviour of this revised model and its outputs are undertaken immediately.

ToR h Evaluate stock status with respect to reference points.

The SSCMSA model allows an integrated approach to estimating stock status and reference points, and therefore there is now consistency between the exploitation rate and abundance reference points. Looking at the time series of exploitation rate and abundance, the revised control rule (Figure 6.14) shows that the crab fishery was overfished in 1998-2004, but by 2009, the exploitation rate was below the $U_{0.75 \times msy}$ reference point and abundance was above the overfishing threshold, but below target abundance. WDS abundance in 2010 was well above this target abundance.

I think that the presentation of this section of the report could be improved to help the reader through the conclusions drawn on stock status in relation to the estimated reference points. Firstly the points on Figures 6.1 and 6.14 should be denoted by the years to which they apply, as this would add to the description in the text of how the stock status has changed over the last 20 years or so in relation to the calculated reference points. Secondly there are two changes that occur between the calculation of the 2005 reference points and those in the 2010 assessment. The panel agreed that it would be very helpful to have an additional figure as follows – Figure 6.1 shows the old method for standardising the survey data and the old reference points and it would therefore be helpful to have an intermediate figure showing the new standardised data for surveys with the old reference points before moving on to Figure 6.14 which shows the new standardised survey data with the newly calculated reference points. This would provide an interpretation of the changes when moving from the 2005 method to the 2010 method.

I note that the population model has been developed as part of a benchmark assessment of the Chesapeake Bay blue crab stock and that there will not be a new assessment each year. The authors recommend that in between these benchmark assessments, the status of the stock with regard to the reference points such as Umsy can be estimated from the empirical estimates of age-0+ female exploitation based on reported female harvests and the abundance of age0+ females in the winter dredge survey calculated using the SSCMSA estimates of q from the model. It is important therefore that any inconsistencies in the model structure are ironed out before the control rule of Figure 6.14 is applied to future empirical estimates of exploitation rate and abundance. This means that the assessment team should reconsider their recommendation about the threshold and target abundance reference points in relation to variable sex-specific exploitation rates, and in particular need to fully examine the sensitivity of the model to the relaxation of the assumption that the WDS is treated as an absolute abundance index.

This will be particularly important because the stock appears to be responding to female only management measures, such as the termination of the winter dredge fishery, which specifically target conservation of the female part of the stock.

Finally whilst not necessarily the role of the assessment team, I would like to see the development of some harvest control rules which kick in if and when reference points are reached or exceeded. At present, fishery managers have a methodology for evaluating the status of the stocks in relation to reference points, but no guidelines on how to respond if those reference points are breached.

Conclusions and Recommendations

The 2011 assessment for the Chesapeake Bay blue crab fishery uses a sex-specific catch multiple survey analysis (SSCMSA) model. This is a major improvement on the 2005 assessment primarily because of the integration of the reference point calculations within the model using per recruit analyses. Previously, the population model and the calculation of reference points were two separate processes which were not necessarily consistent. Two significant concerns with the model were identified. Firstly the formulation of the stock–recruitment relationship implies that it is possible to maintain recruitment even at male populations size of zero, and secondly the model is extremely sensitive to treating the winter dredge survey as a relative abundance index (as opposed to an absolute index of abundance as

assumed in the baseline run). These concerns need to be addressed along with some value judgements about the precautionary nature of the chosen reference points.

The extensive biological knowledge of the species and the wealth of fishery-independent survey data suggest that the fishery is essentially data rich, and there are numerous ways described below where the assessment could be improved in the future.

Specific recommendations

1. Annual variations in growth rate in relation to the allocation of crabs from the surveys to age-0 and age-1+ age classes should be investigated to ensure that assumptions underlying the population model are correct.
2. Sensitivity analysis of the assessment model outputs to the partial recruitment curve generated from the Puckett *et al*. study should be carried out.
3. The population model should be modified to take into account the terminal moult in female crabs replacing the current assumption that all age-1+females contribute to the spawning stock.
4. The potential for sperm limitation in blue crabs should be revisited particularly in the light of recent new management measures to protect the female stock which have caused an increase in the male exploitation rates.
5. Use of the Brownie model with a different parameterisation that includes M and F, or M and catchability and effort to analyse tag-recapture returns would be informative.
6. Natural mortality rate (M) can vary annually, seasonally and with crab size. It is recommended that an analysis is carried out to assess the sensitivity of the model's output to variable M.
7. A spatial component should be added to the population model or at the very least some spatial analysis of the survey data should be undertaken.
8. Recent increases in male exploitation rate and the increased potential for sperm limitation, suggest that male-based reference points might be appropriate, for example, in relation to abundance of age-1+ male crabs.
9. When standardising survey indices using the delta generalised linear model (deltaGLM) method, interaction terms in the second stage of the delta GLM should be considered, especially strata and year, temperature and year for the dredge survey.
10. The winter dredge survey (WDS) is an essential component of the assessment and the evaluation of stock status relative to the reference points and its continuation is strongly recommended.
11. The Virginia (VIMS) trawl survey should be standardized, and better use should be made of the additional four strata.
12. WDS survey results for Virginia and Maryland should be evaluated separately to investigate whether they correlate with trawl survey indices in the two states.
13. Other survey time series, such as the Chesapeake Bay Multispecies Monitoring and Assessment Program, should potentially be included in the assessment.
14. Consideration should be given to the development of an early life history phase recruitment index, e.g. larvae or megalopae abundance.
15. The scale of changes to the time series of catch data following re-calibration to take into account changes in reporting procedures should be independently verified, even if indirectly and/or anecdotally.
16. The assessment team should gain access to, and analyse, the time series of catch data from the Potomac Fisheries Commission.

17. Accurate and complete catch and effort data should be collected from the fishery so that a consistent standardised catch rate dataset can be used in the assessment. The potential existence of personal diaries of daily catch and effort data should be investigated.
18. A baywide survey of recreational catches should be undertaken immediately and at regular intervals to quantify the overall landings from the recreational sector.
19. Some quantitative assessment of under-reporting rates should be obtained.
20. Incidental mortality and discard mortality rates should be ascertained for the pot, trot line and dredge fisheries and survival rates for the peeler/soft crabs held in tanks should be qunatified.
21. Detailed investigation is required as to why the model provides such a poor fit to the sex-specific population sizes.
22. A conceptual model providing a timeline of life history events such as spawning and recruitment should be incorporated.
23. A number of minor changes are required to the formulation of the model. These include defining the initial conditions of the model, decrementing M by time and F by k (kappa) in equations 6.7 and 6.11, incorporating a bias correction and incorporating a compensatory function into the stock-recruitment relationship.
24. The chosen values for the overfished abundance threshold and the target abundance reference point should be re-considered.
25. Consideration should be given to incorporating more biological realism and structure into the model including a size-based approach, shorter time steps, and a spatial component.
26. A more systematic sensitivity analysis should be undertaken to understand how robust the model's output is to the various input parameters.
27. The model should be run with and without the modified time series of catches, when observation error and process error are not estimated simultaneously, and with alternative configurations of the sex-specific stock recruitment relationship.
28. One of the key assumptions of the model is that the abundance of age-1+ crabs from the WDS is an absolute estimate of abundance. As the assessment model is not robust to this assumption, the panel recommended that a sensitivity test is run that uses the raw winter dredge survey indices and that the derivation of the catchability of the winter dredge survey index of abundance should be investigated and possible sources of this difference (both assessment and survey) should be investigated.
29. Additional interpretation should be provided about the two changes that occur between the calculation of the 2005 reference points and those in the 2010 assessment.
30. Harvest control rules should be developed to agree management actions if reference points are reached or exceeded.

Acknowledgements

I would like to thank Tom Miller and all the blue crab assessment team who were refreshingly open and helpful throughout the review, which made it easy for the review panel to understand the strengths and weaknesses of the 2011 assessment. Particular thanks are due to Tom Miller and Mike Wilberg who carried out some additional requested model runs at very short notice. In addition to the assessment team, the review panel benefited greatly from interaction with representatives of the management jurisdictions, and with a wide range of fisheries biologists working in the Blue Crab Advanced Research Consortium, in particular Yoni Zohar who kindly gave us a tour of the blue crab hatchery.

Derek Orner from the NOAA Chesapeake Bay office coordinated all of the review panel meetings and throughout provided a broad perspective on the science, assessment and management of the blue crab population.

References

Aguilar, R., E.G. Johnson, A.H. Hines, P.M. Roberts, M.R. Goodison and M.A. Kramer. 2011. Spatial dynamics of the blue crab (*Callinectes sapidus*) fishery in Chesapeake Bay: understanding post-maturation processes to improve efficacy of marine stock enhancement. Presentation at the Blue Crab Advanced Research Consortium Symposium, Baltimore, Maryland, 28 March 2011.

Brownie, C., D.R. Anderson, K.P. Burnham, and D.S. Robson. 1985. Statistical inference from band recovery data: a handbook. 2nd edition. U.S. Fish and Wildlife Service, Resource Publication 156, Washington D.C.

Fogarty, M.J. and T.J. Miller 2004. Impact of a change in reporting systems in the Maryland blue crab fishery. Fisheries Research **68**: 37-43.

Fonseca, D.B., M.R.J. Sheehy, and P.M.J. Shelton. 2003. Unilateral eyestalk ablation reduces neurolipofuscin accumulation rate in the contra-lateral eyestalk of a crustacean, *Pacifastacus leniusculus*. J. Exp. Mar. Ecol. Biol**. 289**: 277-286.

Haddon, M. 2005. Summary review of the Chesapeake Bay blue crab (*Callinectes sapidus*) 2005 stock assessment.

Harvey, R.H., D.H. Secor and S.-J. Ju. 2008. The use of extractable lipofuscin for age determination of crustaceans: reply to Sheehy. Mar. Ecol. Prog. Ser. **353**: 307-311.

Hoenig, J.M. 1983. Empirical use of longevity data to estimate mortality rates. Fish. Bull. **81:** 898-903.

Johnson, E., A. Hines, H. Soulen, P. Roberts, M. Kramer, M. Goodison and R. Aguilar. 2011. Evidence for large-scale recruitment limitation of blue crabs in Upper Chesapeake Bay. Presentation at the Blue Crab Advanced Research Consortium Symposium, Baltimore, Maryland, 28 March 2011.

Ju, S.-J., D.H. Secor and H.R. Harvey. 1999. Use of extractable lipofuscin for age determination of blue crabs *Callinectes sapidus*. Mar. Ecol. Prog. Ser. **185**: 171-179.

Ju, S.-J., D.H. Secor and H.R. Harvey. 2001. Growth rate variability and LF accumulation rates in the blue crab *Callinectes sapidus*. Mar. Ecol. Prog. Ser. **224**: 197-205.

Kramer, M. A.H. Hines, and E.G. Johnson. 2011. Annual, seasonal and size-dependent patterns of juvenile blue crab, *Callinectes sapidus*, mortality in Upper Chesapeake Bay. Presentation at the Blue Crab Advanced Research Consortium Symposium, Baltimore, Maryland, 28 March 2011.

Lipcius, R.N. 2011. Density dependent exploitation of blue crab in Chesapeake Bay. Assessment Working Paper 2 to Stock assessment of the blue crab in Chesapeake Bay 2011.

Lipcius, R.N. and A. Smith. 2011. Survival, longevity and natural mortality of mature female blue crabs. Assessment Working Paper 1 to Stock assessment of the blue crab in Chesapeake Bay 2011.

Miller, T.J. 2011a. Development and application of simple production models to the Chesapeake Bay blue crab fishery. Assessment Working Paper 3 to Stock assessment of the blue crab in Chesapeake Bay 2011.

Miller, T.J. 2011b. Application and update of a catch, multiple survey model to the Chesapeake Bay blue crab fishery. Assessment Working Paper 4 to Stock assessment of the blue crab in Chesapeake Bay 2011.

Miller, T.J., S.J.D. Martell, D.B. Bunnell, G. Davis, L.A. Fegley, A.F. Sharov, C.F. Bonzek, D.A. Hewitt, J.M. Hoenig and R.N. Lipcius. 2005. Stock assessment of Blue Crab in Chesapeake Bay: 2005. Final Report Ref: [UMCES] CBL 05-077. UMCES Tech. Ser. No. TS-487-05-CBL., University of Maryland Center for Environmental Science Chesapeake Biological Laboratory, Solomons, MD.

Miller, T.J., M.J. Wilberg, A.R. Colton, G.R. Davies, A. Sharov, R.N. Lipcius, G.M. Ralph, E.G. Johnson, and A.G. Kaufman. 2011. Stock assessment of the blue crab in Chesapeake Bay 2011. UMCES Tech. Ser. No. TS-614-11-CBL., University of Maryland Center for Environmental Science Chesapeake Biological Laboratory, Solomons, MD.

Puckett, B.J., D. H. Secor and J-S Ju. 2008. Validation and application of lipofuscin-based age determination for Chesapeake Bay blue crabs (*Callinectes sapidus*). Trans. Am. Fish. Soc. **137**: 1637-1649.

Roberts, P.M., R. Aguilar, E.G. Johnson, A.H. Hines, M.R. Goodison, and M. A. Kramer. 2011 Population and fishery dynamics of blue crab (*Callinectes sapidus*) in a sub-estuary of Chesapeake Bay. Presentation at the Blue Crab Advanced Research Consortium Symposium, Baltimore, Maryland, 28 March 2011.

Rugulo, L.J, K. Knotts, A. Lange, V. Crecco, M. Terceiro, C.F. Bonzek, C. Stagg, R. O'Reilly, and D.S. Vaughan. 1997. Stock assessment of Chesapeake Bay blue crab (*Callinectes sapidus*). National Oceanic and Atmospheric Administration: 267p.

Sheehy, M.R.J. 2008. Questioning the use of biochemical extraction to measure lipofuscin for age determination of crabs: Comment on Ju et al. (1999, 2001). Mar. Ecol. Prog Ser. **353**: 303-306.

Appendix 1 – Bibliography of materials provided for review

Prior to the meeting of the review panel in Baltimore, panel members were provided with password protected access to the University of Maryland, Chesapeake Biological Laboratory website as follows:

http://hjort.cbl.umces.edu/crabs/Assessment.html

This website provided access to the full 2011 assessment document, four associated Working Papers, files containing the catch data, control rule calculations, ADMB code and R scripts and associated data, links to the 2005 assessment and the associated peer review, and other previous assessment papers.

Full assessment report

Miller, T.J., M.J. Wilberg, A.R. Colton, G.R. Davies, A. Sharov, R.N. Lipcius, G.M. Ralph, E.G. Johnson, and A.G. Kaufman. 2011. Stock assessment of the blue crab in Chesapeake Bay 2011. UMCES Tech. Ser. No. TS-614-11-CBL., University of Maryland Center for Environmental Science Chesapeake Biological Laboratory, Solomons, MD.

Working papers

Lipcius, R.N. and A. Smith. 2011. Survival, longevity and natural mortality of mature female blue crabs. Assessment Working Paper 1 to Stock assessment of the blue crab in Chesapeake Bay 2011.

Lipcius, R.N. 2011. Density dependent exploitation of blue crab in Chesapeake Bay. Assessment Working Paper 2 to Stock assessment of the blue crab in Chesapeake Bay 2011.

Miller, T.J. 2011a. Development and application of simple production models to the Chesapeake Bay blue crab fishery. Assessment Working Paper 3 to Stock assessment of the blue crab in Chesapeake Bay 2011.

Miller, T.J. 2011b. Application and update of a catch, multiple survey model to the Chesapeake Bay blue crab fishery. Assessment Working Paper 4 to Stock assessment of the blue crab in Chesapeake Bay 2011.

Appendix 2 CIE Statement of Work – Dr. Julian Addison

Statement of Work for Dr. Julian Addison

External Independent Peer Review by the Center for Independent Experts

Blue Crab Benchmark Stock Assessment - 2010

Scope of Work and CIE Process: The National Marine Fisheries Service's (NMFS) Office of Science and Technology coordinates and manages a contract providing external expertise through the Center for Independent Experts (CIE) to conduct independent peer reviews of NMFS scientific projects. The Statement of Work (SoW) described herein was established by the NMFS Project Contact and Contracting Officer's Technical Representative (COTR), and reviewed by CIE for compliance with their policy for providing independent expertise that can provide impartial and independent peer review without conflicts of interest. CIE reviewers are selected by the CIE Steering Committee and CIE Coordination Team to conduct the independent peer review of NMFS science in compliance the predetermined Terms of Reference (ToRs) of the peer review. Each CIE reviewer is contracted to deliver an independent peer review report to be approved by the CIE Steering Committee and the report is to be formatted with content requirements as specified in **Annex 1**. This SoW describes the work tasks and deliverables of the CIE reviewer for conducting an independent peer review of the following NMFS project. Further information on the CIE process can be obtained from www.ciereviews.org.

Project Description: The blue crab stock has been subject to Baywide stock assessments on two previous occasions. In the years between benchmark assessments, updates on the stock status are provided by the NOAA Chesapeake Bay Office's Chesapeake Bay Stock Assessment Committee. The most recent update concluded that the stock was not overfished and was not then experiencing overfishing. Since the 2005 assessment, the three management jurisdictions have implemented a range of regulatory changes aimed at attaining the target exploitation rate of 46% of the available stock. Thus, it is appropriate that another, Baywide benchmark assessment be conducted. The blue crab resource, specifically for soft and peeler crabs, in Chesapeake Bay has recently been declared a fisheries resource disaster by the Secretary Commerce. In 2009 and 2010, annual updates (not peer-reviewed) have shown slight improvements in the resource. Blue crab is the most important commercial fishery in Chesapeake Bay with annual Baywide landings recently as low as 50 million pounds – roughly 25 million pounds below the long-term average. 2010 predicted landings (if fished at the target exploitation level) could top 100 million pounds. This is obviously a large fluctuation in landings and thus value of the resource to the Bay community.

The first Baywide stock assessment was conducted using a length-based approach to estimate exploitation, and an unweighted average of the four principal fishery-independent surveys to determine abundance. Consequently biological reference points were crude.

In 2001, the technical subcommittee of the Bi-State Blue Crab Advisory Committee (BBCAC) developed a new management framework that relied on exploitation and biomass threshold and target reference points. Threshold reference points were proposed based on a maintaining 10% of the virgin spawning potential and on the lowest observed abundance in

the surveys. A target exploitation rate that would lead to an effective doubling of the spawning stock present in 2001 was also selected. The most recent Baywide benchmark assessment for blue crab in the Chesapeake Bay was conducted in 2005. This assessment critically evaluated and revised estimates of the natural mortality rate, the impact of reporting changes on landings estimates, and spawning potential ratio reference points. The 2005 assessment, using data through 2003, recommended adopting the exploitation fraction, defined as the proportion of crabs available at the beginning of the season that are subsequently harvested, in place of less intuitive measures (F) used in previous assessments. Estimates of exploitation fractions were calculated based on the Baywide winter dredge survey (WDS) and within a modified catch-survey analysis that permitted the use of multiple surveys. The approach used in the 2005 assessment was reviewed by a panel of international scientists with expertise in crustacean fisheries who found that it was a substantial improvement over previous assessments. However, the panel also identified issues to be addressed in future assessments. In particular, the panel recommended exploration of the impact of density-dependent processes in life history traits, improvements to the fishery-independent surveys, particularly with regard to catchability, the possibility of developing a sex-specific assessment model and reference points, and a fuller analysis of the impacts of uncertainty on all aspects of the assessment.

The 2010 assessment and targeted research program is a highly collaborative and integrated program to address specific concerns raised by the international review panel from 2005. The assessment activities are divided into eight specific Terms of Reference (TOR) that were developed based on the review comments received from panel of experts convened to review the 2005 assessment, and from extensive discussion with managers from MDNR, the Potomac River Fisheries Commission and the Virginia Marine Resources Commission, the three relevant management jurisdictions.

NOAA Fisheries is playing a significant role in coordinating disaster assistance to Maryland and Virginia to ensure a sustainable blue crab fishery in Chesapeake Bay. This 2010 Benchmark assessment and research program represents a large investment by NOAA and the state management agencies and should be reviewed internationally.

The Terms of Reference (ToRs) of the peer review are attached in **Annex 2**. The tentative agenda of the panel review meeting is attached in **Annex 3**.

Requirements for CIE Reviewers: Three CIE reviewers shall conduct an impartial and independent peer review in accordance with the SoW and ToRs herein. CIE reviewers shall have working knowledge and recent experience in the application of stock assessment and crustacean fisheries. Each CIE reviewer's duties shall not exceed a maximum of 14 days to complete all work tasks of the peer review described herein.

Location of Peer Review: Each CIE reviewer shall conduct an independent peer review during the panel review meeting scheduled in Baltimore, Maryland during the tentative dates of 29-31 March 2011.

Statement of Tasks: Each CIE reviewers shall complete the following tasks in accordance with the SoW and Schedule of Milestones and Deliverables herein.

Prior to the Peer Review: Upon completion of the CIE reviewer selection by the CIE Steering Committee, the CIE shall provide the CIE reviewer information (full name, title, affiliation, country, address, email) to the COTR, who forwards this information to the NMFS Project Contact no later the date specified in the Schedule of Milestones and Deliverables. The CIE is responsible for providing the SoW and ToRs to the CIE reviewers. The NMFS Project Contact is responsible for providing the CIE reviewers with the background documents, reports, foreign national security clearance, and other information concerning pertinent meeting arrangements. The NMFS Project Contact is also responsible for providing the Chair a copy of the SoW in advance of the panel review meeting. Any changes to the SoW or ToRs must be made through the COTR prior to the commencement of the peer review.

Foreign National Security Clearance: When CIE reviewers participate during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for CIE reviewers who are non-US citizens. For this reason, the CIE reviewers shall provide requested information (e.g., first and last name, contact information, gender, birth date, passport number, country of passport, travel dates, country of citizenship, country of current residence, and home country) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: http://deemedexports.noaa.gov/sponsor.html).

Pre-review Background Documents: Two weeks before the peer review, the NMFS Project Contact will send (by electronic mail or make available at an FTP site) to the CIE reviewers the necessary background information and reports for the peer review. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the CIE Lead Coordinator on where to send documents. CIE reviewers are responsible only for the pre-review documents that are delivered to the reviewer in accordance to the SoW scheduled deadlines specified herein. The CIE reviewers shall read all documents in preparation for the peer review.

Panel Review Meeting: Each CIE reviewer shall conduct the independent peer review in accordance with the SoW and ToRs, and shall not serve in any other role unless specified herein. **Modifications to the SoW and ToRs can not be made during the peer review, and any SoW or ToRs modifications prior to the peer review shall be approved by the COTR and CIE Lead Coordinator.** Each CIE reviewer shall actively participate in a professional and respectful manner as a member of the meeting review panel, and their peer review tasks shall be focused on the ToRs as specified herein. The NMFS Project Contact is responsible for any facility arrangements (e.g., conference room for panel review meetings or teleconference arrangements). The NMFS Project Contact is responsible for ensuring that the Chair understands the contractual role of the CIE reviewers as specified herein. The CIE Lead Coordinator can contact the Project Contact to confirm any peer review arrangements, including the meeting facility arrangements.

Contract Deliverables - Independent CIE Peer Review Reports: Each CIE reviewer shall complete an independent peer review report in accordance with the SoW. Each CIE reviewer shall complete the independent peer review according to required format and content as described in Annex 1. Each CIE reviewer shall complete the independent peer review addressing each ToR as described in Annex 2.

Other Tasks – Contribution to Summary Report: Each CIE reviewer may assist the Chair of the panel review meeting with contributions to the Summary Report, based on the terms of reference of the review. Each CIE reviewer is not required to reach a consensus, and should provide a brief summary of the reviewer's views on the summary of findings and conclusions reached by the review panel in accordance with the ToRs.

Specific Tasks for CIE Reviewers: The following chronological list of tasks shall be completed by each CIE reviewer in a timely manner as specified in the **Schedule of Milestones and Deliverables**.

1) Conduct necessary pre-review preparations, including the review of background material and reports provided by the NMFS Project Contact in advance of the peer review.
2) Participate during the panel review meeting in the Baltimore, Maryland during the tentative dates of 29-31 March 2011, and conduct an independent peer review in accordance with the ToRs (**Annex 2**).
3) No later than 14 April 2011, submit an independent peer review report addressed to the "Center for Independent Experts," and sent to Mr. Manoj Shivlani, CIE Lead Coordinator, via email to shivlanim@bellsouth.net, and Dr. David Sampson, CIE Regional Coordinator, via email to david.sampson@oregonstate.edu. Each CIE report shall be written using the format and content requirements specified in Annex 1, and address each ToR in **Annex 2**.

Schedule of Milestones and Deliverables: CIE shall complete the tasks and deliverables described in this SoW in accordance with the following schedule.

22 February 2011	CIE sends reviewer contact information to the COTR, who then sends this to the NMFS Project Contact
15 March 2011	NMFS Project Contact sends the CIE Reviewers the pre-review documents
29-31 March 2011	Each reviewer participates and conducts an independent peer review during the panel review meeting
14 April 2011	CIE reviewers submit draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator
28 April 2011	CIE submits CIE independent peer review reports to the COTR
5 May 2011	The COTR distributes the final CIE reports to the NMFS Project Contact and regional Center Director

Modifications to the Statement of Work: Requests to modify this SoW must be approved by the Contracting Officer at least 15 working days prior to making any permanent substitutions. The Contracting Officer will notify the COTR within 10 working days after receipt of all required information of the decision on substitutions. The COTR can approve changes to the milestone dates, list of pre-review documents, and ToRs within the SoW as long as the role and ability of the CIE reviewers to complete the deliverable in accordance

with the SoW is not adversely impacted. The SoW and ToRs shall not be changed once the peer review has begun.

Acceptance of Deliverables: Upon review and acceptance of the CIE independent peer review reports by the CIE Lead Coordinator, Regional Coordinator, and Steering Committee, these reports shall be sent to the COTR for final approval as contract deliverables based on compliance with the SoW and ToRs. As specified in the Schedule of Milestones and Deliverables, the CIE shall send via e-mail the contract deliverables (CIE independent peer review reports) to the COTR (William Michaels, via William.Michaels@noaa.gov).

Applicable Performance Standards: The contract is successfully completed when the COTR provides final approval of the contract deliverables. The acceptance of the contract deliverables shall be based on three performance standards:
(1) each CIE report shall completed with the format and content in accordance with **Annex 1**,
(2) each CIE report shall address each ToR as specified in **Annex 2**,
(3) the CIE reports shall be delivered in a timely manner as specified in the schedule of milestones and deliverables.

Distribution of Approved Deliverables: Upon acceptance by the COTR, the CIE Lead Coordinator shall send via e-mail the final CIE reports in *.PDF format to the COTR. The COTR will distribute the CIE reports to the NMFS Project Contact and Center Director.

Support Personnel:

William Michaels, Program Manager, COTR
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
William.Michaels@noaa.gov Phone: 301-713-2363 ext 136

Manoj Shivlani, CIE Lead Coordinator
Northern Taiga Ventures, Inc.
10600 SW 131st Court, Miami, FL 33186
shivlanim@bellsouth.net Phone: 305-383-4229

Roger W. Peretti, Executive Vice President
Northern Taiga Ventures, Inc. (NTVI)
22375 Broderick Drive, Suite 215, Sterling, VA 20166
RPerretti@ntvifederal.com Phone: 571-223-7717

Key Personnel:

NMFS Project Contact:

Derek M. Orner
NOAA Chesapeake Bay Office, 410 Severn Avenue, Suite 107A, Annapolis, MD 21403
Derek.orner@noaa.gov Office: (410) 267-5676 Cell: (410) 570-2268

Annex 1: Format and Contents of CIE Independent Peer Review Report

1. The CIE independent report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether the science reviewed is the best scientific information available.

2. The main body of the reviewer report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs.

   a. Reviewers should describe in their own words the review activities completed during the panel review meeting, including providing a brief summary of findings, of the science, conclusions, and recommendations.

   b. Reviewers should discuss their independent views on each ToR even if these were consistent with those of other panelists, and especially where there were divergent views.

   c. Reviewers should elaborate on any points raised in the Summary Report that they feel might require further clarification.

   d. Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.

   e. The CIE independent report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed, regardless of whether or not they read the summary report. The CIE independent report shall be an independent peer review of each ToRs, and shall not simply repeat the contents of the summary report.

3. The reviewer report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review
   Appendix 2: A copy of the CIE Statement of Work
   Appendix 3: Panel Membership or other pertinent information from the panel review meeting.

Annex 2: Terms of Reference for the Peer Review

Blue Crab Benchmark Stock Assessment - 2010

The stock assessment review has the following eight specific terms of reference:

a) Critically assess and where necessary revise the life history and vital rates of blue crab in the Chesapeake Bay that are relevant to an assessment of the stock.

b) Evaluate and recommend biological reference points for the Chesapeake Bay blue crab population. The potential for implementing sex-specific reference points should be evaluated.

c) Describe and quantify patterns in fishery-independent surveys. Analyses should include an evaluation of the impacts of environmental and abiotic factors on survey catches, to maximize the information content of resultant survey time series.

d) Describe and quantify patterns in catch, effort and survey-based estimates of exploitation by sector and region, including analyses that examine the impacts of reporting changes and trends in CPUE.

e) Develop and implement assessment models for the Chesapeake blue crab fisheries. In particular, models that permit estimates of the trends and status of the crab population and fisheries on a sex-specific basis should be evaluated.

f) Examine density-dependent exploitation patterns derived from survey-based and model-based approaches.

g) Characterize scientific uncertainty with respect to assessment inputs and stock status.

h) Evaluate stock status with respect to reference points.

Annex 3: Tentative Agenda

2010
Blue Crab Stock Assessment Review

Sheraton Baltimore City Center Hotel
101 West Fayette St., Baltimore, MD
March 29-31, 2011

March 29, 2011

12:30 Welcome & Introductions Orner
- Stock Assessment Committee
- Review Panel

12:45 Presentation of the 2010 Blue Crab Stock Assessment Miller

4:00 General / Open Question Period Orner
- Public Comment
- Review Panel

5:30 Adjourn

March 30, 2011

8:30 Term of Reference Review and Discussion

I. Critically assess and where necessary revise the life history and vital rates of blue crab in the Chesapeake Bay that are relevant to an assessment of the stock.

II. Evaluate and recommend biological reference points for the Chesapeake Bay blue crab population. The potential for implementing sex-specific reference points should be evaluated.

III. Describe and quantify patterns in fishery-independent surveys. Analyses should include evaluation of the impacts of environmental and abiotic factors on survey catches, to maximize the information content of resultant survey time-series.

IV. Describe and quantify patterns in catch and effort by sector and region, including analyses that examine the impacts of reporting changes and trends in CPUE.

12:30 Lunch

2010
Blue Crab Stock Assessment Review
Sheraton Baltimore City Center Hotel
101 West Fayette St., Baltimore, MD
March 29-31, 2011

March 30, 2011 (cont.)

1:30 Term of Reference Review and Discussion (continued)

V. Develop and implement assessment models for the Chesapeake blue crab fisheries. In particular, models that permit estimates of the trends and status of the crab population and fisheries on a sex-specific basis should be evaluated.

VI. Examine density-dependent exploitation patterns derives from survey-based and model-based approaches.

VII. Characterize scientific uncertainty with respect to assessment inputs and stock status.

VIII. Evaluate stock status with respect to reference points.

5:15 Adjourn

March 31, 2011

9:00 Review Session *[closed-door]*

- Review Panel to discuss assessment methodologies and develop individual opinions.
- Initiate development of summary documents

12:00 Lunch

1:15 Review Session (continued)
4:30 Adjourn

Appendix 3 - Panel Membership and Review Meeting

The review panel consisted of Dr Julian Addison, an independent fisheries consultant based in France, Dr Cathy Dichmont of CSIRO Marine and Atmospheric Research, Queensland, Australia and Dr Billy Ernst, Department of Oceanography, University of Concepción, Chile. The CIE did not formally appoint a chair of the review panel. The meeting was co-ordinated by Dr Derek Orner of NOAA on behalf of the CIE review team. The panel met formally from 29-31 March at the Sheraton Hotel, Baltimore, Maryland following the agenda outlined in Appendix 2, Annex 3.. The panel met briefly again on 1 April to consider some additional model runs provided by the assessment team, and a summary review was agreed representing the view solely of the CIE panel members.

The Review process worked extremely well. The whole assessment team was extremely helpful and open about the strengths and weaknesses of the new assessment and were very keen to provide any new model runs requested by the review panel. The panel benefitted greatly from having a NOAA Coordinator present throughout the panel meetings to provide broad expertise in the science and management of the fishery. The only minor criticism of the process that I have was that the identity of the panel was not known to panelists or assessment team until the start of the review process.

National Research FLAGSHIPS Wealth from Oceans

CSIRO

Chesapeake Bay Blue Crab Stock Assessment Review 2011

Baltimore, Maryland
29-31 March 2011

Dr Cathy Dichmont
CSIRO Wealth from Oceans Flagship
Ecoscience Precinct, 41 Boggo Road, Dutton Park Qld 4102
AUSTRALIA

www.csiro.au

Representing the Center for Independent Experts

Contents

Executive Summary **4**

Background **5**

Description of Review Activities **6**

Summary of Findings **7**

ToR: Review and Critically assess and where necessary revise the life history and vital rates of blue crab in the Chesapeake Bay that are relevant to an assessment of the stock. **7**

Stock structure 7
Growth rates 7
Moult to maturity 8
Age and size at maturity 8
Mating and spawning periods 8
Fecundity 8
Juveniles 9
Natural mortality 9

ToR: Describe and quantify patterns in fishery-independent surveys. Analyses should include an evaluation of the impacts of environmental and abiotic factors on survey catches, to maximize the information content of resultant survey time series. **10**

Size at age conventions 10
FIS time series 10
Statistical analysis of FIS 10
Virginia juvenile finfish and blue crab trawl survey (since 1955) 11
MD DNR Trawl survey (since 1977) 11
Winter dredge survey (used from 1991/2) 12

ToR: Describe and quantify patterns in catch, effort and survey-based estimates of exploitation by sector and region, including analyses that examine the impacts of reporting changes and trends in CPUE. **13**

Analytical approach to adjusting for reporting changes 13
Effort and catch rate 14
Estimating baywide catches in numbers 14
Recreational harvest 14

ToR: Develop and implement assessment models for the Chesapeake blue crab fisheries. In particular, models that permit estimates of the trends and status of the crab population and fisheries on a sex-specific basis should be evaluated. **15**

ToR: Examine density-dependent exploitation patterns derived from survey-based and model-based approaches. **17**

ToR: Characterize scientific uncertainty with respect to assessment inputs and stock status. **17**

ToR: Evaluate and recommend biological reference points for the Chesapeake Bay blue crab population. The potential for implementing sex-specific reference points should be evaluated. **18**

ToR: Evaluate stock status with respect to reference points. **18**

Changes to the documentation **19**

Conclusions **19**

Future work beyond recommendations **20**

Recommendations with page references **20**

Appendix 1 Documents provided for the review panel **23**
Appendix 2 Review Panel Agenda **24**
Appendix 3 Statement of Work **26**

Executive Summary

The review workshop for the Chesapeake Bay blue crab assessment took place in Baltimore Maryland on March 29-31, 2011. In attendance were review panel members Drs Dichmont, Addison and Ernst, crab stock assessment authors, other scientists involved in the stock assessment, and fishery managers. The review was undertaken in a very co-operative light with reasonable requests for additional work met, including providing more diagnostic plots and a sensitivity test. The panel members were presented with information and reports on basic biological rates and how they were derived, the data inputs to the model, the main assessment, reference points and stock status relative to the reference points. Two reports on additional assessment models were also provided, thereby addressing the sensitivity of the crab assessment to model structure. This information correlated well with the review Terms of Reference, which is appreciated.

Given the three assessments, the main assessment – the sex-specific catch multiple survey analysis (SSCMSA) – is an improvement on past models and the best of the three models submitted for review. The SSCMSA is a two-stage sex disaggregated model that uses three indices of abundance from the surveys, one of which (the winter dredge survey) is treated as absolute. The two stages are year 0 and year 1+ animals. A fundamentally positive development of this model is the integration of the reference point calculations internal to the model using per recruit analyses. This means that assumptions in the assessment can be consistently carried into the reference point calculations. This is a major advance and is highly supported.

This model also differs from its predecessors in that it models male and female populations separately. In light of the recent major changes to the fishery management that have changed the sex ratio of the catch, this is an appropriate advance on previous models.

In order for this integration of incorporating reference points into the model to take place, the model has to incorporate a renewal function, in this case a modified form of the Ricker stock recruitment relationship. This form of the relationship is not plausible in that it assumes very high productivity if females are present, even for a zero male population size i.e. there is no sperm limitation in the population. There is no information on this lower end of the stock-recruitment curve to support such a relationship assumption. A sensitivity test using the traditional approach and a dispensation form was requested but there was not enough time for the assessment team to complete this request given the work required. This is an important aspect of the model that should be further justified given the precautionary approach.

The derivation of the reference point calculations is correctly implemented in the model. The basis of choosing reference points traditionally (at least in other countries' management systems) is not only the domain of the assessment scientists. In this context, the report should provide more choices of reference points, especially the target reference points. Even so, the use of both the Federal reference points system to define the limit reference points and the regional Management Councils' system to define a target reference point is an improvement. However, the assessment team has chosen a limit reference point value that is not in the precautionary part of the potential range as argued.

By international standards, the fishery is considered data rich, especially in terms of the amount and extent of surveys. Some of these data are not included in the assessment - comments regarding these excluded data and which data to assign high priority to are included in the recommendations. Despite the quantum of data, there is uncertainty with respect to the catch and effort data, with the effort data is not used in the model. Although the analyses of adjusted catch data to under-reporting seems reasonable, there is a need to independently verify the assumed values, given that the effect on the catch series is large. Even treating this as a sensitivity test (by using the raw data or a mid range between raw and adjusted) in the model would be an improvement.

A test run during the review highlighted that the assessment model output is extremely sensitive to whether the winter dredge survey is assumed to be a relative or absolute index of abundance. This is in terms of the absolute stock status, reference points and stock status relative to the reference points. Using the relative index model, the likelihood of the model fit is very much lower because the fit to the data is better (despite estimating more parameters). This SCMSA model in its current configuration is therefore not robust to this assumption. This is of great concern, as traditionally surveys are not treated as absolute abundance indices. It is highly recommended that: a) a sensitivity test is run that uses the raw winter dredge survey indices (i.e. not converted to absolute), and b) the derivation of the catchability of the winter dredge survey index of abundance should be checked in detail and possible sources of this difference between input and estimated winter dredge catchabilities (both assessment and survey) investigated.

This assessment is a valid approach and an improvement on the previous assessments and therefore should be adopted as the basis for management advice. However, the justification for treating the winter dredge survey as an absolute index of abundance needs to be substantiated given that the best model fit is achieved by treating the survey as a relative index of abundance (which is the more widely accepted method) and this has important implications to management. This is the highest short-term priority.

Background

Blue crab is the most important commercial fishery in Chesapeake Bay. The blue crab stock has been subject to Baywide stock assessments on two previous occasions. In the years between benchmark assessments, updates on the stock status are provided by the NOAA Chesapeake Bay Office's Chesapeake Bay Stock Assessment Committee.

The first Baywide stock assessment (Rugulo *et al*., 1997) was conducted using a length-based approach to estimate exploitation, and an unweighted average of the four principal fishery-independent surveys to determine abundance. Consequently biological reference points were crude.

In 2001, the technical subcommittee of the Bi-State Blue Crab Advisory Committee (BBCAC) developed a new management framework that relied on exploitation and biomass threshold and target reference points. Threshold reference points were proposed based on maintaining 10% of the virgin spawning potential and on the lowest observed abundance in the surveys. A target exploitation rate that would lead to an effective doubling of the spawning stock present in 2001 was also selected. The most recent Baywide benchmark assessment for blue crab in the Chesapeake Bay was conducted in 2005 (Miller *et al*., 2005). This assessment critically evaluated and revised estimates of the natural mortality rate, the impact of reporting changes on landings estimates, and spawning potential ratio reference points. The 2005

assessment recommended adopting the exploitation fraction, defined as the proportion of crabs available at the beginning of the season that are subsequently harvested, in place of less intuitive measures (F) used in previous assessments. Estimates of exploitation fractions were calculated based on the Baywide winter dredge survey (WDS) and within a modified catch-survey analysis that permitted the use of multiple surveys' data. The approach used in the 2005 assessment was reviewed by a panel of international scientists with expertise in crustacean fisheries who found that it was a substantial improvement over previous assessments (Haddon, 2005). However, the panel also identified issues to be addressed in future assessments. In particular, the panel recommended exploration of the impact of density-dependent processes in life history traits, improvements to the fishery-independent surveys, particularly with regard to catchability, the possibility of developing a sex-specific assessment model and reference points, and a fuller analysis of the impacts of uncertainty on all aspects of the assessment.

Since the 2005 assessment, the three management jurisdictions have implemented a range of regulatory changes aimed at attaining the target exploitation rate of 46% of the available stock. In 2009 the NOAA Chesapeake Bay Office initiated and supported development of another benchmark assessment and that assessment (Miller *et al*., 2011) is the basis of this CIE Peer Review.

Description of Review Activities

The review workshop for the Chesapeake Bay blue crab assessment took place in Baltimore Maryland on March 29-31, 2011. On March 28 there was a Conference by the blue crab advanced research consortium which provided interesting background to the fishery but was not essential to the reviewers. The NOAA Chesapeake Bay Office provided the documents on a website for the Review (Appendix 1). After a request for the code in a form that could be run in ADMB and R, these were provided to the review team via the website.

The objective of the CIE Chesapeake blue crab review was to review and discuss input data, assumptions and parameters to the model, the assessment itself and the biological reference points for the population.

The Terms of Reference were:

a) Critically assess, and where necessary revise, the life history and vital rates of blue crab in the Chesapeake Bay that are relevant to an assessment of the stock.
b) Evaluate and recommend biological reference points for the Chesapeake Bay blue crab population. The potential for implementing sex-specific reference points should be evaluated.
c) Describe and quantify patterns in fishery-independent surveys. Analyses should include an evaluation of the impacts of environmental and abiotic factors on survey catches, to maximize the information content of resultant survey time series.
d) Describe and quantify patterns in catch, effort and survey-based estimates of exploitation by sector and region, including analyses that examine the impacts of reporting changes and trends in CPUE.
e) Develop and implement assessment models for the Chesapeake blue crab fisheries. In particular, models that permit estimates of the trends and status of the crab population and fisheries on a sex-specific basis should be evaluated.
f) Examine density-dependent exploitation patterns derived from survey-based and model-based approaches.
g) Characterize scientific uncertainty with respect to assessment inputs and stock status.

h) Evaluate stock status with respect to reference points.

The meeting was chaired by Dr Derek Orner of NOAA on behalf of the CIE review team, but the summary review was the view of the CIE panel members.

Dr Tom Miller gave a presentation on the first day of the review covering all aspects of the above Terms of Reference (ToR). On the second day, the review panel and assessment team discussed the research against each ToR. Upon request, the team provided additional documentation and information. The assessment team was extremely helpful in this review. The agenda for the review is provided in Appendix 2.

Additional information supplied:

1. Correction of the lagged plots provided in the report
2. Extra residual and predicted versus observed plots
3. Extra run with winter dredge survey as a relative index of abundance, and
4. Extra run using a standard Ricker stock-recruitment function.

Summary of Findings

In order to facilitate a more logical flow for the review, the order of the terms of reference has been slightly changed. This means that comments on the biological reference points are provided after those on the assessment model since these flow from the latter.

ToR: Review and Critically assess and where necessary revise the life history and vital rates of blue crab in the Chesapeake Bay that are relevant to an assessment of the stock.

Stock structure

Based on genetics studies and larval distribution it has been assumed that, from a stock assessment point of view, Chesapeake Bay blue crab is effectively a single stock. This is based on several studies. This assumption seems to be valid given the evidence (which is more than many fisheries making this assumption have had elsewhere).

Growth rates

There are several sources of growth rate information - being lipofuscin analyses, moult process models and stock-recruitment estimates from stock enhancement studies on both larvae, juvenile and small crabs. Growth was shown to be season dependent, which affects the timing of recruitment within the year by 10%. However, growth rates obtained from lipofuscin are not directly or indirectly used in the model other than through inference or body of evidence.

There are three separate processes that have an implicit growth rate:

1. separating the survey size frequency data into age 0 and 1+
2. partial recruitment - a parameter to implicitly link within year indices of abundance to key biological rates, and
3. maximum age and therefore natural mortality.

The survey age classes were derived from a cut off table based on space and time of the survey, however there are very clear modes between age zero and 1+ animals from the graphs provided during the review.

Lipofuscin analysis is used as a method to age crab in Chesapeake Bay. This information is influential. Although there has been some criticism of this method in other parts of the world, it seems appropriately applied here although there could be a question over whether it is a chronometer in oxidative stress areas? In this regard, the ages are provided with a 6 month error range which is therefore appropriate. Furthermore, the sample size of aged crabs is very large.

> **Recommendation 1.** It is unclear whether these different inputs of implicit age are internally consistent. Some evidence of this should be provided.

Moult to maturity

Circumstantial evidence and limited physiological evidence that, for example, the Y organ does not degenerate, point to blue crabs having terminal moult. This assumption, which is influential in the data analysis and model structure, seems appropriate.

Age and size at maturity

Aquaculture studies have shown that females can mature within the first year under ideal conditions. However, the authors argue that given the temperatures during the year and typical megalopial settlement dates, it is unlikely that crabs mature within their first year in Chesapeake Bay. This seems to be an appropriate assumption.

Mating and spawning periods

Female blue crabs are only receptive immediately after moulting. There is intense competition among males during this receptive period. Mating mainly occurs from May to October. In upper Chesapeake Bay none of the upper bay inseminated females will produce broods in the same year as mating whereas an unknown fraction in the lower bay could do so. In the lower Chesapeake Bay, there is evidence that some unknown fraction of females inseminated can release larvae in the same period.

Fecundity

Prager et al. (1990) showed a linear relationship between fecundity and carapace width with r squared of 0.24. Wells (2009) re-examined this relationship for the low-density period and found that there was a significant decrease in size of mature crab from the 1980's to 2005. In recent studies, there is little or no evidence for a linear relationship. Recent, but still preliminary data, show a change in fecundity as a function carapace width occurring from empirical evidence.

> **Recommendation 2.** A possible change in fecundity with respect to size should be monitored in terms of its impact on the assessment.

Dickinson et al. (2006) showed that average sized crabs spawned at the beginning of the season and produce eight clutches within a full 25 week spawning season. Larger animals produce larger clutches, less frequently than smaller animals. Darnell et al. (2009) concluded that the majority of reproductive output of individual females comes from a few initial broods in a female's lifetime.

Recommendation 3. In the assessment model, fecundity is not weighted by age, whereas it is likely to be a consideration and quite a major factor within the assessment where the category age 1+ implies that all animals greater than age 1 are equally fecund even though 2+ animals are unlikely to breed. Splitting the model into three age classes should be considered.

Recommendation 4. A research task for future assessments is to study age related fecundity and other related fecundity patterns for Chesapeake Bay.

Juveniles

Lipcius et al. (2007) showed that the juvenile period is critical for crabs with the state/condition of nursery grounds also being very important. There have been several studies on predation and nursery grounds. Etherington et al. (2003) showed that mortality rates on seagrass were equivalent to emigration rates. There seems to be key lower and upper Chesapeake Bay dynamics and these were not considered within the assessment. Given this and other strong spatial dynamics, it is likely that the assessment should ultimately have a stronger spatial element than it does presently.

Natural mortality

Natural mortality (M) is indirectly calculated using key biological rates, such as maximum age and growth parameters, and environmental factors rather than estimating M directly in the model. This seems appropriate and is the norm internationally.

Originally Rugolo et al. (1997) based his M estimate on a maximum age of eight. Miller at al. (2005) subsequently reviewed estimates of natural mortality for the previous assessment. The values of M used in the present assessment were based on work undertaken by Hewitt et al. (2007), which uses several sources of data to calculate M.

Brownie tag-return models in Lipcius and Smith (Assessment working paper 1) were used to calculate survivorship. However, the extensive tagging data from 2001-2009 (with respect to the assessment) are in essence, only used to calculate maximum age, which is then used to calculate a range of M values. This study therefore calculates M for female crabs that range from 0.6-0.8. The Base Case assessment used 0.9, with sensitivity tests at 0.6 and 1.2. Reference points were tested with the high and low M values as well.

A distribution of M (Fig 2.2 in the Assessment report) gives a range of M from 0.3 - 2.35. The use of 0.9 in the Base Case, rather than using the tagging data values for the Base Case, is due to the fact that the tagging data is of post moult females where mortality is likely to be higher than for the whole population. This assumption appears to be sound. Sensitivity tests to M were undertaken, which is appropriate.

However, the tagging data are underutilised for such an extensive data set. In reality, these data should be incorporated into the model. In addition, the form of the Brownie calculation is fairly basic and could be further improved so that it includes M and F, or M and catchability and effort.

Recommendation 5. In future, use the tagging data in the assessment to estimate female fishing mortality and natural mortality.

Recommendation 6. In the interim, modify the Brownie model to estimate survivorship and natural mortality through a parameterisation that includes M and F, or M and catchability and effort.

ToR: Describe and quantify patterns in fishery-independent surveys. Analyses should include an evaluation of the impacts of environmental and abiotic factors on survey catches, to maximize the information content of resultant survey time series.

Size at age conventions

Previous assessments have used size composition data from fishery independent surveys (FIS) to develop estimates of abundance. Bonzek (pers. comm.) explored the consequences of alternative demarcations of size-at-age vectors.

In this assessment, the size definitions of age classes have not been rigorously or fully evaluated, partly due to inconsistent reporting of size in some surveys. However, figures presented during the review showed reasonably clear modes in the data.

FIS time series

Three FIS are used in the assessment: 49 years of VIMS trawl survey sampling data (taken from the southern portion of the Bay), 28 years of MD trawl survey data (taken in the eastern shores of the Bay) and 16 years of winter dredge survey data (Baywide and treated as an absolute index of abundance).

Data from several other surveys were not used in the assessment because they were either too short or too localised. Potentially the most valuable of these are the Chesapeake Bay Multispecies Monitoring and Assessment Program survey data, even though they have only been gathered since 2003, as it is a Baywide survey. The argument that it is too short for use in the assessment is not supported given that there is a large amount of data already used in the assessment. This means that incorporation of the additional parameters would be supported.

Recommendation 7. Include the Baywide Multispecies Monitoring and Assessment Program in the assessment as a survey index of abundance.

Statistical analysis of FIS

The FIS time series are in the form of count data that are characterised by being zero inflated. Apart from the Virginia trawl survey (which was not standardised), a similar method was used to standardise the FIS indices of abundance. The use of the delta method is appropriate in this instance in that this method models the probability of presence and then models abundance given occurrence. A generalised linear model (GLM) is used to develop standardised indices of abundance. The best fitting model was developed using AIC. The

model considers: Year, Month, Strata, Salinity, Temperature and Depth. For each model fit, variances of the indices were generated using jacknife estimates.

However, given the evidence of strong temporal and spatial structure in the data, it is clear that interaction terms in the GLM should be considered. Although interaction terms cannot be implemented in the package used, this should be modified or remodelled.

> **Recommendation 8.** Incorporate interaction terms in the second stage of the delta GLM, especially with respect to strata and year, temperature and year for the winter dredge survey.

In the delta GLM analysis, there is no link in the assessment report between the final models stated as used in the analyses and the findings shown in Tables 3.4 and 3.5. These were checked during the review and it was shown that they were correctly implemented and the text appears to be incorrect (see section on "Changes to the documentation" for further details).

Virginia juvenile finfish and blue crab trawl survey (since 1955)

There are seven spatial strata within the Virginia trawl surveys, which are undertaken monthly with both a fixed and random survey design. The analysis focused on three principle tributaries, because the spatial extent of the survey changed over time. Therefore only three of the strata are used as an index of abundance in the assessment as it was argued that the four extra strata were added later in the survey design. Furthermore, this index is not standardised. The assessment team rather calculated a geometric mean for each tributary and then took the average of these.

The argument of not including the extra strata is not supported given that there are years where a mechanism of comparison was included in the Virginia trawl survey expansion coverage. Either a standardisation process could have combined this data set or these extra strata could have been added as an additional survey.

> **Recommendation 9.** Include in the assessment model either using the four other Virginia trawl survey strata as an additional index of abundance or (better option) analyse the Virginia trawl survey and standardise into a single index.

MD DNR Trawl survey (since 1977)

This survey has inconsistent coverage both spatially and temporally. The survey is conducted from May to November and uses fixed sites. From 1989, size measurements were taken. The best model for age 0 abundance indices were developed using both design factors and temperature.

In Fig 3.9 of the assessment document, the two peaks in the modelled index have no corresponding observed data points. During the review, it was shown that the observations are outside the upper limit of the axes in the plots. This should be corrected in the document.

Winter dredge survey (used from 1991/2)

This is a key dataset as it is used in the assessment model as an absolute index of abundance. This is unusual given that most modellers are unable to accurately calculate survey catchability relative to the whole population (therefore treating survey indices as being a relative index of abundance). The survey is a stratified random design and then a subsequent adaptive design is used where additional sampling occurs in high-density areas.

The survey data have extensive spatial and temporal components, which appear to have not been used. See "Statistical analysis of FIS" section above.

The dredge efficiency figures are obtained from depletion analyses, which are undertaken in high-density areas for practical reasons. Absolute abundance is extrapolated to the whole area using a GIS estimated area. It should be noted that each jurisdiction uses different survey gear and different survey designs. The survey gear catch efficiency is estimated for each year, but not with respect to spatial distribution changes.

> **Recommendation 10.** Conduct more detailed analyses of the dredge survey catchability estimates, especially with regard to availability or else undertake a spatial model of the survey.
>
> **Recommendation 11.** Consider applying consistent methods across jurisdictions.
>
> **Recommendation 12.** The two-stage component of the dredge survey is an extremely valuable component of the analysis and should be continued.
>
> **Recommendation 13.** The dredge survey is essential and should be continued.

During the review, a sensitivity test was undertaken in which the assessment was modified so as to treat the winter dredge survey index as a relative index of abundance. This causes substantial changes to the model outputs – the model estimates much lower dredge survey catchability, fits the winter dredge survey index better and estimates of fishing mortality are higher than in the Base Case. Furthermore, this test has much better likelihood values than the Base Case, which treats the winter dredge survey as an absolute index of abundance. These results show that the assessment is extremely sensitive to this assumption, in terms of absolute stock status, reference points and stock status relative to reference points (see Appendix 3). This model is therefore not robust to this assumption, which is of great concern given that indices of abundance are traditionally not treated as an absolute index of abundance. In the review, the panel was not given detailed information about the winter dredge survey in order to clearly analyse the validity of this assumption.

> **Recommendation 14.** Run a sensitivity test that uses the raw standardised winter dredge survey indices (not converted to absolute).
>
> **Recommendation 15.** Review the derivation of the absolute index of abundance in detail and investigate possible sources of the difference between the absolute and relative model runs (both in terms of the assessment and the survey). This is the highest short-term priority.

ToR: Describe and quantify patterns in catch, effort and survey-based estimates of exploitation by sector and region, including analyses that examine the impacts of reporting changes and trends in CPUE.

Several gear types are used to catch blue crab – there is a recreational and several commercial sectors. Two jurisdictions, Maryland and Virginia, manage this fishery within their jurisdiction. The Potomac River Fisheries Commission manages the Potomac River fisheries. Each jurisdiction has made several (and often different) management decisions in the history of this fishery. There are three sources of commercial fisheries data, being from Virginia, Maryland and the Potomac River Fisheries Commission. The latter data are daily from the Potomac River, but are not used or analysed as a potential source of commercial CPUE in the assessment.

> **Recommendation 16.** Use the daily Potomac data in the assessment, which seem to be very rich and useful.

Analytical approach to adjusting for reporting changes

There have been significant changes to the reporting systems, however the most influential are argued to be those that occurred during 1981. A classic Box-Jenkins time series model is used to analyse these changes. Tests of stationarity in some cases showed that the analyses had to be undertaken on the time series that are in the form of differences. The analysis assumes that current data are more accurate and landings from this period are the most reliable. The effects of this analysis on the Virginia and Maryland estimated commercial landings are significant.

Although the method appears reasonable, the very large effect on the results should be validated. Experience has shown that the time of change is more readily agreed upon the quantum. There are several ways of addressing these issues. These should be prioritised based on ease of implementation.

> **Recommendation 17.** Independently substantiate these adjusted changes to the catch data through interviews of dealers/fishers or investigate whether there is a discrepancy within the spatial landings data as the fishery effort moves over time.

> **Recommendation 18.** Undertake model sensitivity tests with unadjusted catch data and an intermediate value of the reporting change factor.

> **Recommendation 19.** Estimate the reporting change factor (or intervention parameter) internal to the model (best option). If the model is unable to estimate this parameter, then an alternative would be to include the reporting change parameter (with variances) as an input value within the model so sensitivity tests of this parameter can be undertaken and the error in this variable could be included internal to the model.

The different gear types in the fishery are not considered, yet these can be substantially different in location and affect the size and type of crab caught.

Recommendation 20. Investigate the potential of dividing the catch data into gear type or at least undertake a gear analysis.

Effort and catch rate

Neither effort or catch rate data are included in assessment. This is due to the fact that data appear to be inconsistent over time and between jurisdictions.

Recommendation 21. Investigation should be carried out to establish whether it is possible to produce a standardised catch rate dataset that can be used in the assessment, as the index only needs to be representative rather than complete. Also, investigate the use of effort data within the assessment.

Recommendation 22. Given the value of this fishery in the region, there should be more emphasis placed on collecting accurate and complete catch and effort data that are well aligned with the needs of the assessment.

The soft and peeler crab catch is based on landings of animals that have survived being kept in a shedding tank, whereas the animals that died during the process is not included as a component of the catch.

Recommendation 23. Either record the catch of the soft and peeler crab at the point of landing or record the length of time the crabs are kept in the tank and estimate the survival rate of these crabs.

Estimating baywide catches in numbers

The catches are usually reported in bushels and need to be converted to numbers. In recent years, the method of converting to numbers has been to calculate the average carapace width of crabs in the population on an annual and sex specific basis based on the Maryland and Virginia trawl surveys, and then apply to a regression of weight to carapace length based on data from Maryland trawl data pooled between 1994 to 2004. It is unclear (where available) why the distribution itself was not used rather than the average.

Recommendation 24. The conversion of catch in bushels to numbers should use the mean weight from the catch for each year both for the past and the future. This means that the mean size in the catch by year is required.

Recreational harvest

Recreational catch is estimated to be 5.3 to 8.5% of the commercial catch based on surveys in 2001 and 2002. Localised tagging studies pointed to recreational catch being much higher than 8% based on very large return rates of tags from the recreational sector. The uncertainty is whether these studies represent the whole bay. These surveys are the only Baywide surveys and are well analysed and are therefore the only reliable data of value to the assessment.

Recommendation 25. Undertake a baywide recreational survey as the resource is increasing and recreational effort may be increasing. This is a priority.

Recommendation 26. Undertake sensitivity tests for larger recreational catches in the assessment.

ToR: Develop and implement assessment models for the Chesapeake blue crab fisheries. In particular, models that permit estimates of the trends and status of the crab population and fisheries on a sex-specific basis should be evaluated.

The present assessment being reviewed is a two-stage sex disaggregated model that uses three indices of abundance from the surveys, one of which is treated as absolute. The two stages are year 0 and year 1+ animals. About 144 parameters (for the Base Case) are estimated in ADMB. A fundamental positive development of this model is the integration of the reference point calculations internal to the model using per recruit analyses. This means that assumptions in the assessment can be consistently carried into the reference point calculations. This is a major improvement and highly supported. In order for this integration to take place the model has to incorporate a renewal function, in this case a modified form of the Ricker stock recruitment relationship.

This model also differs from its predecessors in that it tracks males and females separately. In light of major recent changes to the fishery management that have changed the sex ratio of the catch, this is an appropriate development on previous models.

There is some controversy concerning the chosen formulation of the stock-recruitment relationship since it implies that the population will remain highly productive while there are any females even at zero male population size i.e. there is not sperm limitation in the population. This form of the relationship is not plausible. There is no information on the lower end of the stock-recruitment curve to assume such a relationship. A sensitivity test of using the traditional approach was requested and provided for comparison and, although the results shown did not completely follow through to the stock status relative to the reference points, it shows very similar results to the Base Case (note these were preliminary runs). This is not unexpected as the model is implemented using a reasonable sex ratio based upon observed data. However, the reality is that this model optimises yield-per-recruit at a sex ratio of zero males. This aspect should be clearly stated in the documentation to highlight that this assessment model should not be used to optimise the sex ratio using the assessment's per recruit analysis. This is an important aspect of the model that should be further justified given the precautionary approach.

Recommendation 27. Undertake full sensitivity tests of the stock-recruitment relationship (following this into the reference point calculations) and also develop a function that captures the principle of what is presently formulated but does not optimise at zero males i.e. incorporate a compensatory function in the stock-recruitment relationship.

In the documentation there was some confusion with regard to how the different indices of abundance were linked to crab biology.

Recommendation 28. Develop a conceptual model which would provide a much better description of the biological timeline and how this connects to the assessment.

In many of the equations in the assessment document (e.g. equation 7,11) both M and F are decremented by an effort related parameter. It would be more appropriate to index M by time of year and F by effort.

Recommendation 29. The decrementation in, for example, equations 7 and 11 etc. should index M with time of year and F by effort.

The partial recruitment parameter in equation 11 captures growth but should also capture selectivity. This confuses what are two separate processes.

Recommendation 30. Split the partial recruitment parameter (that ties selectivity and growth) to include gear specific selectivity. The model should also overtly include selectivity of the gear (including how it changed over time).

The initial conditions of the model use two parameters to set up the numbers matrix. This is not mentioned in the assessment documentation. It is more correct to set up the model such that the initial conditions match the assessment model and would therefore require four (not two) parameters. This is not likely to have a large effect on the model, though.

Recommendation 31. For consistency, set the initial conditions correctly allied with the assessment implementation i.e. from 2 to 4 parameters.

The effective sample size iterative method was used for the two trawl surveys, but given the extensive nature of the dredge survey index, only the calculated variance is used. This seems correctly implemented in the trawl survey cases. If possible, a similar process should be undertaken for the dredge survey.

Recommendation 32. Include the calculation of effective sample size for the winter dredge survey in the likelihood of the assessment.

There should be a bias correction in the stock-recruitment relationship so that the estimated parameters and the per-recruit analysis are consistent with international practice and would therefore not be incorrectly used in other studies.

Recommendation 33. Include the bias correction in the stock-recruitment relationship.

The model consistently fits the trawl survey data poorly. This is in part due to the contradictory nature of the data, but also due to the fact that this assessment is set up as both a process and observation error model. The model tends to treat the poor data fits as

observation error and therefore does not capture the high and low recruitment years. There would be some benefit in running a test where the model is formulated as either process or observation error driven.

> **Recommendation 34.** Run sensitivity tests where the model is formulated as either process or observation error driven.

ToR: Examine density-dependent exploitation patterns derived from survey-based and model-based approaches.

Lipcius (Assessment working paper 2) undertook an in-depth analysis of density-dependent exploitation of blue crabs in Chesapeake Bay. The paper presented empirical evidence that exploitation rate varies inversely with population size and that exploitation rate is depensatory. It is unclear how or whether this work is carried into the assessment other than through attempting to choose more precautionary reference points (see below). The use of the values in Lipcius could be considered as a sensitivity test.

ToR: Characterize scientific uncertainty with respect to assessment inputs and stock status.

There were several sensitivity tests undertaken with respect to a) M, b) the degree of partial recruitment, c) estimating male M while fixing the female M to the Base Case value, d) estimating the sex ratio and e) estimating the partial recruitment. It is laudable that the sensitivity table includes the sensitivity of the assessment parameters, but also the stock status relative to the reference points. However, in reality the sensitivity tests should be much more extensive than those provided. Several recommendations are included throughout the text in this document and will not be repeated here.

There are several survey indices within the model and the effect of each individual index is not tested within the model. This should be adopted as standard practice.

> **Recommendation 35.** Undertake a sensitivity test of the influence of the different indices as well as implement the dredge survey as a relative index.

There should have been more sensitivity tests and the documentation should provide confidence intervals on outputs.

> **Recommendation 36.** Run sensitivity tests of different input data and parameters and its effects on reference points. Consider systematic approaches such as FAST or a designed experiment.

In the text, there were some model fit statistics provided, but this was inadequate to fully characterise uncertainty. For example, many of the plots of the indices and their fits did not include residuals. Also, a table of the actual output parameter estimates and variances was not provided.

> **Recommendation 37.** Provide more detailed fit statistics like q-q plots and residual plots for all indices.

Two additional assessment models were provided for review or information – the previous assessment model, but updated with recent data and a production model. As a test of model uncertainty, this is highly unusual and laudable. Given the differences in model structure between the main assessment model and these models, the uncertainty in stock status relative to the reference points cannot be tested. Of the three models tested, the SSCMSA, which is the model used to derive the 2010 reference points, is the most appropriate.

ToR: Evaluate and recommend biological reference points for the Chesapeake Bay blue crab population. The potential for implementing sex-specific reference points should be evaluated.

ToR: Evaluate stock status with respect to reference points.

These two terms of references have been addressed together as they tend to be discussed as a unit.

The derivation of the reference point calculations is correctly implemented in the model even though there is an unusual stock-recruitment relationship. The harvest control rule is based on a mixture of model output to set the values for the reference points, whereas the indices over time are calculated from empirical data. The reason for this formulation is that the assessment is not undertaken each year and as such there was a need to develop a harvest control rule that can be used in the intermediate years.

The exploitation limit reference point (overfishing) is based on the fact that the optimal yield against female age 0+ exploitation fraction is insensitive to the input sex ratio value. The exploitation fraction target reference point is 75% of the limit reference, which is partly based on the method applied by the regional Mid-Atlantic and New England management councils.

The biological target index is based on the absolute abundance of female 1+ crabs from the winter dredge survey. However, in this case the female age 1+ (Spawning) abundance relative to catch in the per-recruit analysis is sensitive to the sex ratio assumed. The assessment team chose a spawning size at the lower end of the range from the sex ratio results of 0.6 to 0.8. The ultimate limit reference point is therefore 50% of the spawning size at maximum sustainable yield, similar to that used in the Federal system.

The basis of choosing reference points traditionally (at least in other countries) is not only the domain of the assessment scientists. In this context, the report should provide more choices of reference points especially the target reference points. Even so, the use of both the Federal reference points system to define the limit reference points and the regional Management Councils' system to define a target reference point is a step forward.

It is important that when different sensitivity tests are undertaken that they are consistent throughout including the empirical indicators. As an example, the test that assumed the winter dredge survey was a relative index of abundance required that the empirical values of the spawning index had to be adjusted appropriately to incorporate the estimated catchability value.

> **Recommendation 38.** Provide more choices of reference points rather than providing single values for each.

The assessment team has chosen a limit reference point value that is not from the precautionary end of the potential range as argued.

Recommendation 39. Provide the option for more precautionary limit reference points.

Changes to the documentation

The documentation describing the model is well explained and set out mathematically, however the following modifications should be incorporated prior to its final delivery:

a. A conceptual model which would provide a much better description of the biological timeline and how this connects to the assessment. This component should also specify the assumptions.
b. More detailed description of previous tests undertaken, provision of estimated parameter values and variances, and uncertainty around model outputs.
c. Sensitivity tests undertaken during the review.
d. The bias correction should be included in the text.
e. Correct the lagged plots as updated during the review.
f. Correct Figure 3.9 to include all the data points.
g. Correct the text when describing the factors chosen in the deltaGLM analyses as presently these are not the best model as highlighted in the tables (Tables 3.4 and 3.5) – the latter was used based on responses to this question during the review.
h. Clarification of the initial conditions as this is not presently documented.
i. Full documentation of the non-precautionary nature of the stock-recruitment relationship particularly at its extremes.
j. The references to the Figures from Figure 6.8 onwards are not always correct which makes confusing reading.
k. The reference to the simple production model Section 6.3.1 is incorrect.
l. Data points on the reference point figures (6.1 and 6.14) should be annotated with year labels, as this would link better with the text.
m. Correct the legend of Figure 6.12.
n. Create a reference point figure which uses the empirical data as analysed by the present 2010 method while using the 2005 reference points i.e. the document would contain three reference points figures as follows: a) old method for standardising the survey data and old reference points, b) new standardised data with old reference points and c) new standardised data with new reference points. The documentation should include an interpretation of the changes when moving from the 2005 to the 2010 method.

Recommendation 40. The assessment document should be updated to include various corrections and clarifications as described in the review report.

Conclusions

This assessment is a valid approach and an improvement on the previous assessments and therefore should be adopted as the basis for management advice. However, the justification for treating the winter dredge survey as an absolute index of abundance needs to be

substantiated given that the best model fit is achieved by treating the survey as a relative index of abundance (which is the more widely accepted method) and this has important implications to management. This is the highest short-term priority.

Future work beyond recommendations

This assessment would benefit from being size-based with a shorter time step – this would internalise the growth rate assumptions, allow in-season changes to be accommodated and also incorporate different gear types. Adding some spatial structure to the model should also be considered.

There are clear spatial-temporal changes in the data that have not been fully captured in the assessment. Other crab studies have shown the value of detailed spatial and temporal mining of the data. There would be value in undertaking a detailed and cohesive analysis of spatial and temporal dynamics of different population components on different surveys and their relationship with abiotic factors.

Recommendations with page references

Recommendation 1. It is unclear whether these different inputs of implicit age are internally consistent. Some evidence of this should be provided. 8

Recommendation 2. A possible change in fecundity with respect to size should be monitored in terms of its impact on the assessment. 8

Recommendation 3. In the assessment model, fecundity is not weighted by age, whereas it is likely to be a consideration and quite a major factor within the assessment where the category age 1+ implies that all animals greater than age 1 are equally fecund even though 2+ animals are unlikely to breed. Splitting the model into three age classes should be considered. 9

Recommendation 4. A research task for future assessments is to study age related fecundity and other related fecundity patterns for Chesapeake Bay. 9

Recommendation 5. In future, use the tagging data in the assessment to estimate female fishing mortality and natural mortality. 10

Recommendation 6. In the interim, modify the Brownie model to estimate survivorship and natural mortality through a parameterisation that includes M and F, or M and catchability and effort. 10

Recommendation 7. Include the Baywide Multispecies Monitoring and Assessment Program in the assessment as a survey index of abundance. 10

Recommendation 8. Incorporate interaction terms in the second stage of the delta GLM, especially with respect to strata and year, temperature and year for the winter dredge survey. 11

Recommendation 9. Include in the assessment model either using the four other Virginia trawl survey strata as an additional index of abundance or (better option) analyse the Virginia trawl survey and standardise into a single index. 11

Recommendation 10. Conduct more detailed analyses of the dredge survey catchability estimates, especially with regard to availability or else undertake a spatial model of the survey. 12

Recommendation 11. Consider applying consistent methods across jurisdictions. 12

Recommendation 12. The two-stage component of the dredge survey is an extremely valuable component of the analysis and should be continued. 12

Recommendation 13. The dredge survey is essential and should be continued. 12

Recommendation 14. Run a sensitivity test that uses the raw standardised winter dredge survey indices (not converted to absolute). 12

Recommendation 15. Review the derivation of the absolute index of abundance in detail and investigate possible sources of the difference between the absolute and relative model runs (both in terms of the assessment and the survey). This is the highest short-term priority. 12

Recommendation 16. Use the daily Potomac data in the assessment, which seem to be very rich and useful. 13

Recommendation 17. Independently substantiate these adjusted changes to the catch data through interviews of dealers/fishers or investigate whether there is a discrepancy within the spatial landings data as the fishery effort moves over time. 13

Recommendation 18. Undertake model sensitivity tests with unadjusted catch data and an intermediate value of the reporting change factor. 13

Recommendation 19. Estimate the reporting change factor (or intervention parameter) internal to the model (best option). If the model is unable to estimate this parameter, then an alternative would be to include the reporting change parameter (with variances) as an input value within the model so sensitivity tests of this parameter can be undertaken and the error in this variable could be included internal to the model. 13

Recommendation 20. Investigate the potential of dividing the catch data into gear type or at least undertake a gear analysis. 14

Recommendation 21. Investigation should be carried out to establish whether it is possible to produce a standardised catch rate dataset that can be used in the assessment, as the index only needs to be representative rather than complete. Also, investigate the use of effort data within the assessment. 14

Recommendation 22. Given the value of this fishery in the region, there should be more emphasis placed on collecting accurate and complete catch and effort data that are well aligned with the needs of the assessment. 14

Recommendation 23. Either record the catch of the soft and peeler crab at the point of landing or record the length of time the crabs are kept in the tank and estimate the survival rate of these crabs. 14

Recommendation 24. The conversion of catch in weight (bushels) to numbers should use the mean weight from the catch for each year both for the past and the future. This means that the mean size in the catch by year is required. 14

Recommendation 25. Undertake a baywide recreational survey as the resource is increasing and recreational effort may be increasing. This is a priority. 15

Recommendation 26. Undertake sensitivity tests for larger recreational catches in the assessment. 15

Recommendation 27. Undertake full sensitivity tests of the stock-recruitment relationship (following this into the reference point calculations) and also develop a function that captures the principle of what is presently formulated but does not optimise at zero males i.e. incorporate a compensatory function in the stock-recruitment relationship. 15

Recommendation 28. Develop a conceptual model which would provide a much better description of the biological timeline and how this connects to the assessment. 16

Recommendation 29. The decrementation in, for example, equations 7 and 11 etc. should index M with time of year and F by effort. 16

Recommendation 30. Split the partial recruitment parameter (that ties selectivity and growth) to include gear specific selectivity. The model should also overtly include selectivity of the gear (including how it changed over time). 16

Recommendation 31. For consistency, set the initial conditions correctly allied with the assessment implementation i.e. from 2 to 4 parameters. 16

Recommendation 32. Include the calculation of effective sample size for the winter dredge survey in the likelihood of the assessment. 16

Recommendation 33. Include the bias correction in the stock-recruitment relationship. .. 16

Recommendation 34. There would be some benefit in running sensitivity tests where the model is formulated as either process or observation error driven. 17

Recommendation 35. Undertake a sensitivity test of the influence of the different indices as well as implement the dredge survey as a relative index. 17

Recommendation 36. Run sensitivity tests of different input data and parameters and its effects on reference points. Consider systematic approaches such as FAST or a designed experiment. 17

Recommendation 37. Provide more detailed fit statistics like q-q plots and residual plots for all indices. 17

Recommendation 38. Provide more choices of reference points rather than providing single values for each. 18

Recommendation 39. Provide the option for more precautionary limit reference points. . 19

Recommendation 40. The assessment document should be updated to include various corrections and clarifications as described in the review report. 19

Appendix 1 Documents provided for the review panel

Prior to the meeting of the review panel in Baltimore, the full assessment document was provided to the panel –

Miller, T.J., M.J. Wilberg, A.R. Colton, G.R. Davies, A. Sharov, R.N. Lipcius, G.M. Ralph, E.G. Johnson, and A.G. Kaufman. 2011. Stock assessment of the blue crab in Chesapeake Bay 2011. UMCES Tech. Ser. No. TS-614-11-CBL., University of Maryland Center for Environmental Science Chesapeake Biological Laboratory, Solomons, MD.

Files containing the catch data, control rule calculations, ADMB code and R scripts were provided to the review panel.

In addition four working papers were provided:

- Lipcius, R.N. and A. Smith. 2011. Survival, longevity and natural mortality of mature female blue crabs. Assessment Working Paper 1 to Stock assessment of the blue crab in Chesapeake Bay 2011.

- Lipcius, R.N. 2011. Density dependent exploitation of blue crab in Chesapeake Bay. Assessment Working Paper 2 to Stock assessment of the blue crab in Chesapeake Bay 2011.

- Miller, T.J. 2011. Development and application of simple production models to the Chesapeake Bay blue crab fishery. Assessment Working Paper 3 to Stock assessment of the blue crab in Chesapeake Bay 2011.

- Miller, T.J. 2011. Application and update of a catch, multiple survey model to the Chesapeake Bay blue crab fishery. Assessment Working Paper 4 to Stock assessment of the blue crab in Chesapeake Bay 2011.

Access was also given to past assessments.

Appendix 2 Review Panel Agenda

2010

Blue Crab Stock Assessment Review

Sheraton Baltimore City Center Hotel

101 West Fayette St.

Baltimore, MD

March 29-31, 2011

March 29, 2011

12:30	Welcome & Introductions	Orner
	- Stock Assessment Committee	
	- Review Panel	
12:45	Presentation of the 2010 Blue Crab Stock Assessment	Miller
4:00	General / Open Question Period	Orner
	- Public Comment	
	- Review Panel	
5:30	Adjourn	

March 30, 2011

8:30 Term of Reference Review and Discussion

I. Critically assess and where necessary revise the life history and vital rates of blue crab in the Chesapeake Bay that are relevant to an assessment of the stock.

II. Evaluate and recommend biological reference points for the Chesapeake Bay blue crab population. The potential for implementing sex-specific reference points should be evaluated.

III. Describe and quantify patterns in fishery-independent surveys. Analyses should include evaluation of the impacts of environmental and abiotic factors on survey catches, to maximize the information content of resultant survey time-series.

IV. Describe and quantify patterns in catch and effort by sector and region, including analyses that examine the impacts of reporting changes and trends in CPUE.

12:30 Lunch

1:30 Term of Reference Review and Discussion (continued)

V. Develop and implement assessment models for the Chesapeake blue crab fisheries. In particular, models that permit estimates of the trends and status of the crab population and fisheries on a sex specific basis should be evaluated.

VI. Examine density-dependent exploitation patterns derives from survey-based and model-based approaches.

VII. Characterize scientific uncertainty with respect to assessment inputs and stock status.

VIII. Evaluate stock status with respect to reference points.

5:15 Adjourn

March 31, 2011

9:00 Review Session *[closed-door]*

- Review Panel to discuss assessment methodologies and develop individual opinions.
- Initiate development of summary documents

12:00 Lunch

1:15 Review Session (continued)

4:30 Adjourn

Appendix 3 Statement of Work

External Independent Peer Review by the Center for Independent Experts

Blue Crab Benchmark Stock Assessment - 2010

Scope of Work and CIE Process: The National Marine Fisheries Service's (NMFS) Office of Science and Technology coordinates and manages a contract providing external expertise through the Center for Independent Experts (CIE) to conduct independent peer reviews of NMFS scientific projects. The Statement of Work (SoW) described herein was established by the NMFS Project Contact and Contracting Officer's Technical Representative (COTR), and reviewed by CIE for compliance with their policy for providing independent expertise that can provide impartial and independent peer review without conflicts of interest. CIE reviewers are selected by the CIE Steering Committee and CIE Coordination Team to conduct the independent peer review of NMFS science in compliance the predetermined Terms of Reference (ToRs) of the peer review. Each CIE reviewer is contracted to deliver an independent peer review report to be approved by the CIE Steering Committee and the report is to be formatted with content requirements as specified in **Annex 1**. This SoW describes the work tasks and deliverables of the CIE reviewer for conducting an independent peer review of the following NMFS project. Further information on the CIE process can be obtained from www.ciereviews.org.

Project Description: The blue crab stock has been subject to Baywide stock assessments on two previous occasions. In the years between benchmark assessments, updates on the stock status are provided by the NOAA Chesapeake Bay Office's Chesapeake Bay Stock Assessment Committee. The most recent update concluded that the stock was not overfished and was not then experiencing overfishing. Since the 2005 assessment, the three management jurisdictions have implemented a range of regulatory changes aimed at attaining the target exploitation rate of 46% of the available stock. Thus, it is appropriate that another, Baywide benchmark assessment be conducted. The blue crab resource, specifically for soft and peeler crabs, in Chesapeake Bay has recently been declared a fisheries resource disaster by the Secretary Commerce. In 2009 and 2010, annual updates (not peer-reviewed) have shown slight improvements in the resource. Blue crab is the most important commercial fishery in Chesapeake Bay with annual Baywide landings recently as low as 50 million pounds – roughly 25 million pounds below the long-term average. 2010 predicted landings (if fished at the target exploitation level) could top 100 million pounds. This is obviously a large fluctuation in landings and thus value of the resource to the Bay community.

The first Baywide stock assessment was conducted using a length-based approach to estimate exploitation, and an unweighted average of the four principal fishery-independent surveys to determine abundance. Consequently biological reference points were crude.

In 2001, the technical subcommittee of the Bi-State Blue Crab Advisory Committee (BBCAC) developed a new management framework that relied on exploitation and biomass threshold and target reference points. Threshold reference points were proposed based on a maintaining 10% of the virgin spawning potential and on the lowest observed abundance in the surveys. A target exploitation rate that would lead to an effective doubling of the

spawning stock present in 2001 was also selected. The most recent Baywide benchmark assessment for blue crab in the Chesapeake Bay was conducted in 2005. This assessment critically evaluated and revised estimates of the natural mortality rate, the impact of reporting changes on landings estimates, and spawning potential ratio reference points. The 2005 assessment, using data through 2003, recommended adopting the exploitation fraction, defined as the proportion of crabs available at the beginning of the season that are subsequently harvested, in place of less intuitive measures (F) used in previous assessments. Estimates of exploitation fractions were calculated based on the Baywide winter dredge survey (WDS) and within a modified catch-survey analysis that permitted the use of multiple surveys. The approach used in the 2005 assessment was reviewed by a panel of international scientists with expertise in crustacean fisheries who found that it was a substantial improvement over previous assessments. However, the panel also identified issues to be addressed in future assessments. In particular, the panel recommended exploration of the impact of density-dependent processes in life history traits, improvements to the fishery-independent surveys, particularly with regard to catchability, the possibility of developing a sex-specific assessment model and reference points, and a fuller analysis of the impacts of uncertainty on all aspects of the assessment.

The 2010 assessment and targeted research program is a highly collaborative and integrated program to address specific concerns raised by the international review panel from 2005.

The assessment activities are divided into eight specific Terms of Reference (TOR) that were developed based on the review comments received from panel of experts convened to review the 2005 assessment, and from extensive discussion with managers from MDNR, the Potomac River Fisheries Commission and the Virginia Marine Resources Commission, the three relevant management jurisdictions.

NOAA Fisheries is playing a significant role in coordinating disaster assistance to Maryland and Virginia to ensure a sustainable blue crab fishery in Chesapeake Bay. This 2010 Benchmark assessment and research program represents a large investment by NOAA and the state management agencies and should be reviewed internationally.

The Terms of Reference (ToRs) of the peer review are attached in **Annex 2**. The tentative agenda of the panel review meeting is attached in **Annex 3**.

Requirements for CIE Reviewers: Three CIE reviewers shall conduct an impartial and independent peer review in accordance with the SoW and ToRs herein. CIE reviewers shall have working knowledge and recent experience in the application of stock assessment and crustacean fisheries. Each CIE reviewer's duties shall not exceed a maximum of 14 days to complete all work tasks of the peer review described herein.

Location of Peer Review: Each CIE reviewer shall conduct an independent peer review during the panel review meeting scheduled in Baltimore, Maryland during the tentative dates of 29-31 March 2011.

Statement of Tasks: Each CIE reviewers shall complete the following tasks in accordance with the SoW and Schedule of Milestones and Deliverables herein.

<u>Prior to the Peer Review</u>: Upon completion of the CIE reviewer selection by the CIE Steering Committee, the CIE shall provide the CIE reviewer information (full name, title, affiliation, country, address, email) to the COTR, who forwards this information to the NMFS Project Contact no later the date specified in the Schedule of Milestones and Deliverables. The CIE is

responsible for providing the SoW and ToRs to the CIE reviewers. The NMFS Project Contact is responsible for providing the CIE reviewers with the background documents, reports, foreign national security clearance, and other information concerning pertinent meeting arrangements. The NMFS Project Contact is also responsible for providing the Chair a copy of the SoW in advance of the panel review meeting. Any changes to the SoW or ToRs must be made through the COTR prior to the commencement of the peer review.

Foreign National Security Clearance: When CIE reviewers participate during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for CIE reviewers who are non-US citizens. For this reason, the CIE reviewers shall provide requested information (e.g., first and last name, contact information, gender, birth date, passport number, country of passport, travel dates, country of citizenship, country of current residence, and home country) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: http://deemedexports.noaa.gov/sponsor.html).

Pre-review Background Documents: Two weeks before the peer review, the NMFS Project Contact will send (by electronic mail or make available at an FTP site) to the CIE reviewers the necessary background information and reports for the peer review. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the CIE Lead Coordinator on where to send documents. CIE reviewers are responsible only for the pre-review documents that are delivered to the reviewer in accordance to the SoW scheduled deadlines specified herein. The CIE reviewers shall read all documents in preparation for the peer review.

Panel Review Meeting: Each CIE reviewer shall conduct the independent peer review in accordance with the SoW and ToRs, and shall not serve in any other role unless specified herein. **Modifications to the SoW and ToRs can not be made during the peer review, and any SoW or ToRs modifications prior to the peer review shall be approved by the COTR and CIE Lead Coordinator.** Each CIE reviewer shall actively participate in a professional and respectful manner as a member of the meeting review panel, and their peer review tasks shall be focused on the ToRs as specified herein. The NMFS Project Contact is responsible for any facility arrangements (e.g., conference room for panel review meetings or teleconference arrangements). The NMFS Project Contact is responsible for ensuring that the Chair understands the contractual role of the CIE reviewers as specified herein. The CIE Lead Coordinator can contact the Project Contact to confirm any peer review arrangements, including the meeting facility arrangements.

Contract Deliverables - Independent CIE Peer Review Reports: Each CIE reviewer shall complete an independent peer review report in accordance with the SoW. Each CIE reviewer shall complete the independent peer review according to required format and content as described in Annex 1. Each CIE reviewer shall complete the independent peer review addressing each ToR as described in Annex 2.

Other Tasks – Contribution to Summary Report: Each CIE reviewer may assist the Chair of the panel review meeting with contributions to the Summary Report, based on the terms of reference of the review. Each CIE reviewer is not required to reach a consensus, and should provide a brief summary of the reviewer's views on the summary of findings and conclusions reached by the review panel in accordance with the ToRs.

Specific Tasks for CIE Reviewers: The following chronological list of tasks shall be completed by each CIE reviewer in a timely manner as specified in the **Schedule of Milestones and Deliverables**.

1) Conduct necessary pre-review preparations, including the review of background material and reports provided by the NMFS Project Contact in advance of the peer review.
2) Participate during the panel review meeting in the Baltimore, Maryland during the tentative dates of 29-31 March 2011, and conduct an independent peer review in accordance with the ToRs (**Annex 2**).
3) No later than 14 April 2011, submit an independent peer review report addressed to the "Center for Independent Experts," and sent to Mr. Manoj Shivlani, CIE Lead Coordinator, via email to shivlanim@bellsouth.net, and Dr. David Sampson, CIE Regional Coordinator, via email to david.sampson@oregonstate.edu. Each CIE report shall be written using the format and content requirements specified in Annex 1, and address each ToR in **Annex 2**.

Schedule of Milestones and Deliverables: CIE shall complete the tasks and deliverables described in this SoW in accordance with the following schedule.

22 February 2011	CIE sends reviewer contact information to the COTR, who then sends this to the NMFS Project Contact
15 March 2011	NMFS Project Contact sends the CIE Reviewers the pre-review documents
29-31 March 2011	Each reviewer participates and conducts an independent peer review during the panel review meeting
14 April 2011	CIE reviewers submit draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator
28 April 2011	CIE submits CIE independent peer review reports to the COTR
5 May 2011	The COTR distributes the final CIE reports to the NMFS Project Contact and regional Center Director

Modifications to the Statement of Work: Requests to modify this SoW must be approved by the Contracting Officer at least 15 working days prior to making any permanent substitutions. The Contracting Officer will notify the COTR within 10 working days after receipt of all required information of the decision on substitutions. The COTR can approve changes to the milestone dates, list of pre-review documents, and ToRs within the SoW as long as the role and ability of the CIE reviewers to complete the deliverable in accordance with the SoW is not adversely impacted. The SoW and ToRs shall not be changed once the peer review has begun.

Acceptance of Deliverables: Upon review and acceptance of the CIE independent peer review reports by the CIE Lead Coordinator, Regional Coordinator, and Steering Committee, these reports shall be sent to the COTR for final approval as contract deliverables based on compliance with the SoW and ToRs. As specified in the Schedule of Milestones and

Deliverables, the CIE shall send via e-mail the contract deliverables (CIE independent peer review reports) to the COTR (William Michaels, via William.Michaels@noaa.gov).

Applicable Performance Standards: The contract is successfully completed when the COTR provides final approval of the contract deliverables. The acceptance of the contract deliverables shall be based on three performance standards:

(1) each CIE report shall completed with the format and content in accordance with **Annex 1**,

(2) each CIE report shall address each ToR as specified in **Annex 2**,

(3) the CIE reports shall be delivered in a timely manner as specified in the schedule of milestones and deliverables.

Distribution of Approved Deliverables: Upon acceptance by the COTR, the CIE Lead Coordinator shall send via e-mail the final CIE reports in *.PDF format to the COTR. The COTR will distribute the CIE reports to the NMFS Project Contact and Center Director.

Support Personnel:

William Michaels, Program Manager, COTR
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
William.Michaels@noaa.gov Phone: 301-713-2363 ext 136

Manoj Shivlani, CIE Lead Coordinator
Northern Taiga Ventures, Inc.
10600 SW 131st Court, Miami, FL 33186
shivlanim@bellsouth.net Phone: 305-383-4229

Roger W. Peretti, Executive Vice President
Northern Taiga Ventures, Inc. (NTVI)
22375 Broderick Drive, Suite 215, Sterling, VA 20166
RPerretti@ntvifederal.com Phone: 571-223-7717

Key Personnel:

NMFS Project Contact:

Derek M. Orner
NOAA Chesapeake Bay Office,
410 Severn Avenue, Suite 107A,
Annapolis, MD 21403
Derek.orner@noaa.gov
Office: (410) 267-5676 Cell: (410) 570-2268

Annex 1: Format and Contents of CIE Independent Peer Review Report

1. The CIE independent report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether the science reviewed is the best scientific information available.

2. The main body of the reviewer report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs.

   a. Reviewers should describe in their own words the review activities completed during the panel review meeting, including providing a brief summary of findings, of the science, conclusions, and recommendations.

   b. Reviewers should discuss their independent views on each ToR even if these were consistent with those of other panelists, and especially where there were divergent views.

   c. Reviewers should elaborate on any points raised in the Summary Report that they feel might require further clarification.

   d. Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.

   e. The CIE independent report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed, regardless of whether or not they read the summary report. The CIE independent report shall be an independent peer review of each ToRs, and shall not simply repeat the contents of the summary report.

3. The reviewer report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review
   Appendix 2: A copy of the CIE Statement of Work
   Appendix 3: Panel Membership or other pertinent information from the panel review meeting.

Annex 2: Terms of Reference for the Peer Review

Blue Crab Benchmark Stock Assessment - 2010

The stock assessment review has the following eight specific terms of reference:

a) Critically assess and where necessary revise the life history and vital rates of blue crab in the Chesapeake Bay that are relevant to an assessment of the stock.

b) Evaluate and recommend biological reference points for the Chesapeake Bay blue crab population. The potential for implementing sex-specific reference points should be evaluated.

c) Describe and quantify patterns in fishery-independent surveys. Analyses should include an evaluation of the impacts of environmental and abiotic factors on survey catches, to maximize the information content of resultant survey time series.

d) Describe and quantify patterns in catch, effort and survey-based estimates of exploitation by sector and region, including analyses that examine the impacts of reporting changes and trends in CPUE.

e) Develop and implement assessment models for the Chesapeake blue crab fisheries. In particular, models that permit estimates of the trends and status of the crab population and fisheries on a sex-specific basis should be evaluated.

f) Examine density-dependent exploitation patterns derived from survey-based and model-based approaches.

g) Characterize scientific uncertainty with respect to assessment inputs and stock status.

h) Evaluate stock status with respect to reference points.

Blue Crab Stock Assessment Review
2010

March 2011
Baltimore, Maryland

Billy Ernst

Department of Oceanography

University of Concepción

Chile

Executive summary

This report is a review of the 2010 blue crab benchmark assessment in Chesapeake Bay. This assessment is based on a new sex and two-stage structured model. The demographic model follows the dynamics of abundance of year 0 and year 1+ crabs and has a modified Ricker stock recruitment model as a renewal function. It allows for adult males and females to contribute to compensatory mechanisms affecting recruitment at larger population sizes and primarily adult females to affect recruitment at low population levels. The model is fitted to four main sources of information, including overall catch of Chesapeake Bay blue crab, Baywide winter dredge survey, Maryland DNR trawl survey and Virginia juvenile finfish and blue crab trawl survey. A sensitivity analysis to model assumptions is presented in the assessment document. Biological reference points are defined for stock abundance and exploitation rates.

The 2010 assessment model is a major improvement to its predecessors (2005 and 1997) and is a valid approach. Biological reference points are calculated internally and correctly implemented in the model and are consistent with underlying model assumptions. This is a major advance with respect to previous models. Keeping track of males and females separately is appropriate considering that current management has affected the sex ratio.

The main points of concern with respect to this assessment are:

- The assessment model results are not robust to treating the winter dredge survey as a relative index of abundance. If q is different from one the overall model fit improves substantially and is a much easier assumption to support.
- Model structure configuration implicitly assumes that female fecundity does not change with age. The impact of this on the reference point calculation is unknown.
- Stock recruitment function is not plausible with respect to mature male abundance.

An extended list of short term and future recommendations are presented in this document. It is recommended that the short-term list be considered before drafting the final report.

Recommendations

Short term

The document

1. There is an appropriate description and mathematical representation of the stock assessment model in the main document; nevertheless it is necessary that the authors include some modifications in the final document:

a. A conceptual model should be included, which would provide a better description of the components, biological timeline and how this connects to the assessment, especially with regard to survey data. Main model assumptions should also be addressed.

b. The stock recruitment function should include the bias correction factor to avoid misinterpretation of recruitment parameters estimates.

c. Initial conditions of the population dynamic model should be clearly specified in the document.

d. Appropriate documentation should be provided of the non-precautionary nature of the stock-recruitment function, particularly at low population stock size.

e. A better description is needed of the sensitivity tests presented in the document, including estimated parameter values and standard errors, likelihood values associated to each data component, and uncertainty around the main model outputs.

f. Sensitivity tests undertaken during the review should be described.

g. Data points on the reference point figures (6.1 and 6.14) should be annotated with the years, as this would link better with the text.

h. A composite reference point figure should be provided, including: a) the 2005 method for standardizing the survey data and the 2005 reference points, b) the 2010 standardized data with the 2005 reference points and c) the 2010 standardized data with the 2010 reference points. This figure should be accompanied with an interpretation of the changes when moving from the 2005 to the 2010 assessment.

2. In many ways this fishery can be considered data rich and from all available data the winter dredge survey is the most important piece of information. It is used in the assessment and the evaluation of stock status relative to the reference points. It is important that this survey is continued.

3. The structure of model (Age-0/Age-1+ and sex) seems appropriate considering the nature of the historical data. It also allows one to keep track of meaningful population statistics (spawning abundance by sex), which is of special relevance under recent changes to the management regime in the fishery.

4. The calculation of reference points is incorporated within the assessment model. This is a major advantage over previous models, because it assures consistency between statistics.

5. One inconvenience of collapsing the entire adult age structure into one age group is losing track of female fecundity at age. The implicit assumption is that female fecundity remains constant, which is probably rather unlikely to happen. Additionally it is unclear how this would influence the spawning stock size at MSY calculations.

6. In the assessment report there was an important effort to assess model structure uncertainty (results of two additional models were presented). Despite the intrinsic value of this exercise, a direct comparison of reference points between the models is not possible, because female spawning stock is not available for the other models. Nevertheless of the three assessment models provided, the best model is the SSCMSA.

7. On the other hand the parameter uncertainty of the proposed model was not fully characterized through the documented sensitivity tests, nor was the robustness of reference points estimates adequately presented. For example, test runs during the review highlight that the assessment model output is extremely sensitive to treating the winter dredge survey as a relative index of abundance. The fit was much better than the base case presented in the report. This is of great concern, as traditionally the default assumption is to treat surveys as a relative index of abundance. It is recommended that:

a) A sensitivity test is run using the raw winter dredge survey indices.

b) The entire calculation of survey catchability should be checked in detail and possible sources of difference should be investigated.

8. The reference point calculations seem appropriate based on the structure of the new model. The authors followed the Federal reference points system to define the limit reference points and the New England and Mid-Atlantic Fishery Management Councils' system to define target reference points. This seems appropriate, but is in the end a policy decision. With respect to stock based reference points there is a value judgment about how precautionary the overfished limit and target reference point needs to be. Different values arise using different male to female F ratios, and coincidentally the one chosen in the report is not the most precautionary one.

Future work

9. This assessment lacks a catch per unit effort index of abundance, an index that captures the condition of the stock from the fisheries perspective. Future assessments should investigate the feasibility of generating a standardized catch rate index. This would ultimately be of importance in future economic analyses.

10. The substantial difference between the model and estimated catchability coefficient of the winter dredge survey requires immediate attention. This is a high priority task.

11. The spatial component seems to play an important role in the population dynamics of different ontogenetic stages of Chesapeake Bay blue crab. A detailed data-driven analysis (combining all of the pieces of information) of spatial and temporal dynamics of different stock components should be undertaken. Environmental covariates should also be factored in.

12. More work on the natural mortality parameter estimation should be undertaken using the available mark recapture data. Alternative Brownie model parameterization should be considered in the analysis, to explicitly model natural and instantaneous fishing mortality rates. Ultimately, the tagging data could be directly incorporated into a more integrated stock assessment estimation model.

13. Only a nominal index of abundance was constructed from the Virginia trawl survey data during this assessment. Index standardization should be undertaken. A second index of abundance can be potentially developed based on the winter portion of the Virginia trawl survey, which could provide information on the female spawning stock.

14. Interaction terms were omitted in the standardization of survey indices. It is recommended to develop additional models with interaction terms in the second stage of the delta GLM analysis.

15. In recent years some areas in the Chesapeake Bay seem to be experiencing higher recreational fishing mortality than reported in this analysis. It is recommended that a new baywide recreational survey be undertaken.

16. Despite the data-rich nature of this assessment, there is a major need for accurate and complete catch and effort information. Some of these factors are presented in the following list: (a) the conversion of the catch in weight to numbers should be based on measured average weight in the catch, rather than using a constant factor. This means that mean size in the catch by year is required; (b) a better and more direct estimate of soft and peeler crab mortality should be developed for the assessment; (c) the adjustment of past changes in catch reporting requires some independent verification. The catch data is influential to the assessment results; (d) Anecdotal information suggests that incidental winter dredge and summer pot fishery mortality might be important. It is important to quantify these additional sources of mortality.

17. Additional model runs should be undertaken to assess the robustness of the results to: (a) using the modified/original time series of catch; (b) estimating the catchability coefficient of winter dredge survey for adult crab; (c) using different pieces of information; (d) not estimating observation and process error simultaneously; (e) to free up the assessment model to allow for a better fit of extreme values. A sensitivity test might be organized in a systematic and comprehensive way, using appropriate techniques (i.e. FAST, designed experiments).

18. Include additional sources of information into the assessment such as Baywide Multispecies Monitoring and Assessment Program and Potomac daily catch data.

19. Adjust the reporting changes in the catch internal to the assessment model and input the intervention parameter so that this acknowledges some of the uncertainty.

20. Incorporate formally the different fisheries into the assessment. In addition reparameterize the partial recruitment parameter into formal selectivity and growth components.

21. The initial conditions of the model should have 4 as opposed to 2 parameters. If initial conditions stay in a 2-parameter configuration, the implicit assumption for the other 2 parameters should be consistent with the rest of the model configuration.

22. In equation 7 and 11, kappa should be multiplied by only F and M should be scaled by time.

23. Incorporate some kind of compensatory component into the stock recruitment relationship to avoid non-precautionary behavior of the model (i.e., depleting male population, while maintaining high recruitment rate).

24. A more complex version of the model might include: (a) size structure; (b) monthly time step to allow for the evaluation of in-season regulations; and (c) spatial structure. The model development could be frameworked into a MSE type of configuration.

Conclusions

The 2010 assessment model is a major improvement to its predecessors (2005 and 1997) and is a valid approach.

This fishery is data-rich based on fisheries independent information, but data-poor from fisheries dependent perspective. More effort needs to be devoted to improving catch and effort data.

The modified Ricker stock recruitment function is non-precautionary and that needs to be addressed in the document.

The assessment models results are not robust to departures in the assumption of winter dredge survey being an absolute index of abundance.

Background

NOAA Chesapeake Bay Office through the Center for Independent Experts (CIE) requested an independent review for the 2010 Chesapeake Bay blue crab stock assessment. This review included three fisheries independent surveys, fisheries dependent information, the 2010 stock assessment model and the reference point calculation of blue crab (*Callinectes sapidus*) in Chesapeake Bay.

Blue crab is the most important commercial fishery and an icon in the Chesapeake Bay region. The blue crab stock has been subject to Baywide stock assessment on two previous occasions. The first Baywide stock assessment was developed by Rugolo et al. (1997). It concluded that the stock was moderately to fully exploited and at average levels of abundance. The second blue crab stock assessment was reported by Miller et al. (2005) and consisted of a Catch-Survey-Analysis (Collie and Sissenwine, 1983) that used fisheries dependent and independent data. Reference points were calculated using individual based yield per recruit analysis (Bunnell and Miller 2005) and stock status estimates were compared to these values. Based on these reference points, the assessment concluded that exploitation rates in the fishery were too high. Since 2005, the status of the blue crab stock has been updated annually and its status determined relative to the reference points. In 2010 a third Baywide assessment was conducted through a highly collaborative and integrated program to address specific concerns raised by the international CIE review panel from 2005. The 2010 assessment activities are divided into eight specific Terms of Reference (TOR), and are the base for the current CIE Peer Review.

Description of review activities

Before and during the meeting

Documentation

Two weeks before the staff meeting at the Sheraton Baltimore City Center Hotel in Baltimore, several papers and software from the official "*Assessment of blue crab in Chesapeake Bay"* review webpage were uploaded:

http://hjort.cbl.umces.edu/crabs/Assessment.html

This website link was timely provided by Mr. Derek Orner.

The list of documents was the following:

- Executive Summary.
- Full Assessment.
- Working paper 1 - Survival analysis.

- Working paper 2 - Depensation analysis.
- Working paper 3 - Production modeling.
- Working paper 4 - CMSA model.
- Research Recommendations.

Previous Assessments

- 2005 Assessment.
- Limits and Targets (2001).
- Miller and Houde (1999).
- Rugolo et al. (1997).

The list of software was the following:

- Catch Data (EXCEL document).
- Control Rule Calculations (EXCEL document).
- ADMB code and input file.
- R scripts.

During the meeting and upon request, the team provided the following additional documentation, information and tests:

1. Correction of the lagged plots
2. Extra residual and predicted versus observed plots
3. An additional model run assuming that the winter dredge survey is a relative index of abundance.

Before the meeting

Panel members were invited by Mr. Derek Orner to attend on Monday the 28th the Blue Crab Advanced Research Consortium (BCARC) presentations at the National Shellfish Association's Annual Meeting in Baltimore. Interesting and updated information on blue crab ecology was presented all day long, providing some background information on this species.

Review Activities

The review workshop was conducted at the Sheraton Baltimore City Center Hotel in Baltimore, Maryland over three days: Tuesday the 29th to Thursday the 31st of March 2011. An extra morning of work (Friday April 2011) was necessary to wait for extra runs results and complete the Summary Report from CIE Review Panel Members.

The assessment team was composed of the following four groups and most of them participated in the review:

- Thomas Miller, Michael Wilberg and A.R. Colton from Chesapeake Biological Laboratory, University of Maryland Center for Environmental Science, Solomons, MD.
- G. R. Davis, A. F. Sharov. Fisheries Service, Maryland Department of Natural Resources, Annapolis, MD
- R. N. Lipcius, G. Ralph. Virginia Institute of Marine Science, Gloucester Point, VA
- E. G. Johnson, A. G. Kaufman, Smithsonian Environmental Research Center, Edgewater, MD

Other people present throughout the review

- Robert O'Reilly. Virginia Marine Resource Commission.
- Derek Orner. National Oceanic and Atmospheric Administration Chesapeake Bay Office.

The review panel was constituted by:

- Dr. Catherine Dichmont (CSIRO, Marine and Atmospheric Research, Queensland, Australia).
- Dr. Billy Ernst (UDEC, Concepción, Chile).
- Dr. Julian Addison (independent fisheries consultant, France).

The meeting was coordinated by Mr. Derek Orner of NOAA on behalf of the CIE review team, but the summary review was the view solely of the CIE panel members.

Dr Tom Miller gave a presentation on the first day of the review covering all aspects of the above Terms of Reference (ToR). On the second day, the review panel and assessment team discussed the research with the assessment team against each ToR. The assessment team was extremely helpful in this review.

Thursday and Friday morning were devoted by the review panel members for discussion and preparation of the CIE review panel member summary report. The timetable of presentations and discussions is presented in Annex 3.

Findings and recommendations

ToR A: Critically assess and where necessary revise the life history and vital rates of blue crab in the Chesapeake Bay that are relevant to an assessment of the stock.

Stock structure

The Chesapeake Bay blue crab is assessed as an independent stock, separated from Delaware and southern populations. A definitive statement about spatial population structure from a genetic standpoint is still lacking, but larval distribution studies provide evidence of a quasi-discrete stock and movement of benthic stages is restricted to the Chesapeake Bay or adjacent areas in the estuary. Based on available information it seems appropriate to treat the Chesapeake Bay blue crab benthic population as a separate stock. Random stock exchange with adjacent areas during early life stages is possible in the assessment model through the recruitment residuals parameters of the stock-recruitment function.

Conceptual model

This species, as many other crustaceans, has a complex life history with extended and massive female reproductive migrations, rapid growth and maturation, female terminal molt, crab over wintering and high natural mortality. Probably some of these processes are affected by gradients and fluctuations in environmental conditions.

There is a need for better documentation of a comprehensive and formal conceptual model that encompass the life history, spatial dynamics and timing of different benthic ontogenetic stages of blue crab in Chesapeake Bay. This conceptual framework might help during the development of scientific surveys, interpretation of the data and the development of stock assessment models of different levels of complexity. The conceptual model will become very important for defining parsimonious operating and estimation models in a future management strategy evaluation (MSE) framework (Smith et al. 1999).

The different trawl and dredge surveys constitute a valuable source of information that provides both spatial and temporal contrast to the data. This spatially explicit database should be analyzed thoroughly to reveal spatial pattern and spatial dynamics of this species and its relationship with the environment. This can be entirely a data driven approach such as the one reported by Ernst et al. (2005) for snow crab in the Eastern Bering Sea.

Fecundity

The relationship between fecundity and size is important, as indicated in the document Prager et al. (1990), but the key issue is how female fecundity changes with age.

The 2010 stock assessment model keeps track of Age-0 and Age-1+ males and females. Assuming Age-1+ is a good proxy of mature females, the model lumps all adult females

in one category, regardless of their age. This could be appropriate if fecundity does not vary with age, but is probably not the case for this species. As with other terminal molt crustaceans (e.g. snow crab), fecundity probably drops as female senesce (Orensanz et al 2005). Female blue crabs mate only once, right after the last molting event. Anecdotal information generated during the review discussions indicates that most of the sperm load is used during the first spawning season, through multiple spawning batches, which probably makes this iteroparous species to effectively reproduce as a quasi-semelparous one. This factor can be of major importance in spawning size calculations and its impact on reference point calculation is probably relevant.

It would be valuable to construct an empirically based clutch fullness index of egg-berried females to assess time dependent changes in female fecundity. This is an extra covariate that needs to be measured during the surveys (Orensanz et al 2005) and is a valuable tool to identify sperm limitation.

Mating

Under the current model configuration the effective sex ratio at mating becomes an important variable to assess, because it affects reference point determination. Experimental laboratory work might be valuable for defining biologically precautionary sex ratios at mating. It is important to mention that sex ratios should be computed at scales that are meaningful to the mating process, especially for this species that has extensive ontogenetic sex/age dependent latitudinal migrations.

Growth

A growth function is not explicitly part of this assessment model, because length or biomass information is not used as hard data in the assessment. The only link between age and size is at the moment or parsing the survey data into the two-stage categories. These assignations are based on size cut-points that were derived from historical modal decomposition analysis applied upon survey based size frequency data (this was explained to us during the presentations). Some figures that were shown to us showed very distinct modes, which probably justify the chosen cut-points. From the provided data it is hard to judge how much uncertainty is involved in the age assignation (especially for different times of the year and sexes) and it would be valuable to put forward a quantitative analysis that characterizes some of this uncertainty.

Age information derived from Ju et al. (1999) and Puckett et al. 2008 is also used to condition the partial recruitment parameter.

Natural mortality

It is important to acknowledge that since the first assessment a substantial amount of effort has been allocated to assess natural mortality. This effort is greater than expected for a regular stock assessment. Estimates of natural mortality were thoroughly reviewed in a previous assessment (Miller et al 2005). Indirect methods for estimating M were developed using growth parameters, age at maturity, longevity and temperature. Natural

mortality values from these methods ranged from 0.3 to 2.35. The distribution is centered at 1.1, a value that is substantially higher than the natural mortality estimate used in the 1997 assessment. Despite potential problems associated with indirect approaches described by Pascual and Iribarne (1993), these methods provide parameter estimates that integrate various life-stages throughout the ontogeny of blue crab.

Several medium to long-term initiatives are providing valuable mark-recapture information on terminal molted female blue crab in Chesapeake Bay. This is a major breakthrough because it allows for direct estimates of natural mortality from this stock.

The survival analysis paper by Lipcius and Smith (2011) used a Brownie model to estimate yearly survival and recovery rates. Model selection was used to define the optimal saturation level of the model, being the one with time dependent survival and recovery rates. Unfortunately M was not estimated directly from the Brownie model, but using again longevity information from the mark-recapture data. Two venues of model development aimed at improving the estimation of natural mortality are proposed:

- Use an alternative parameterization of the Brownie model, one which includes explicitly the catch equation for harvest rate calculation (Frusher and Hoenig 2001), with or without fishing effort information.
- Integrate the mark recapture data directly into the assessment model (Maunder 2001).

Despite the method used for estimating M with current mark-recapture data, it is important to remember that it reflects survival rates associated with terminal molted females. It is recommended that in a medium time framework an additional mark-recapture experiment be implemented for males and Age-0 individuals. Males do not have terminal molt and natural mortality rates might be different than females. This would require a different kind of tag (i.e. T-bars).

ToR B: Evaluate and recommend biological reference points for the Chesapeake Bay blue crab population. The potential for implementing sex-specific reference points should be evaluated.

One important advantage and improvement of the current assessment model over its predecessors is the possibility of calculating directly biologically meaningful and consistent reference points. The derivation of reference point calculations is adequately implemented in the model and considers a sex-specific formulation of the stock-recruitment relationship. Once the biological reference points are estimated, limit reference points are calculated using the Federal reference point system. This seems to be totally adequate. Because of the nature of the stock recruitment model and the possibility to harvest males and females differently, several yield curves are produced, depending on F_{males}: $F_{females}$ ratios. Implicit biological reference points are various, and there is a value judgment about how precautionary the overfished limit and target

reference point needs to be. It would be more convenient to provide the stakeholders with a matrix that would lay out possible combinations of fishing effort ratios and implicit reference points. The limit reference points that were chosen by the assessment team do not represent the more precautionary ones.

Target reference points, on the other side are also based on biological reference points from the model, but use the Mid-Atlantic and New England Fishery Management Councils' system to define them. This is appropriate, but is again a value judgment and there are several others that could be considered. The choice is a policy decision.

Potential relevant factors that might affect biological reference points that deserve further attention are:

- Dredge survey index being treated as a relative index of abundance for Age-1+ crabs. This is of major concern.
- Factoring in fecundity into the spawning stock calculation and having fecundity vary by age (as is probably the case).

ToR C: Describe and quantify patterns in fishery-independent surveys. Analyses should include evaluation of the impacts of environmental and abiotic factors on survey catches, to maximize the information content of resultant survey time-series.

The strongest component of the 2010 Chesapeake Bay blue crab assessment is the fisheries independent data. Three different surveys were considered for this stock assessment; (a) The Virginia Institute of Marine Sciences (VIMS) trawl survey (1968 – 2009), (b) Maryland trawl survey (1977-2009) and (c) The Winter Dredge Survey (WDS) (1989-2009).

VIMS juvenile finfish and blue crab trawl survey

This annual survey consists typically of 60 stations, organized in 7 strata. Historically random and fixed assignments of stations have been employed. In this assessment only the spring portion of this trawl survey was used, because migration is less important in that part of the year. To eliminate biases of different coverage areas over time, only three main tributaries were considered in the assessment (James, Cork, Rappahannock rivers). The VIMS trawl survey time series was not standardized and only nominal catch per tows values are available. Correction of gear catchability was introduced directly to the data to compensate for two conspicuous changes in the gear. Age-1+ female and male abundance show very similar patterns, a highly variable period up to 1990 and low levels of variability and abundance after that. Correlation analysis between Age-0(t) and Age-1+(t+1) crab were presented. There is not a high correlation between these two metrics. Given the spatial dynamics of these species there should not necessarily be such a correlation for females, because an important fraction of Age-1+ females have migrated from the upper Bay. Until the migration patterns and rates by sex are better understood, such correlations are probably not very meaningful.

MD DNR trawl survey

This trawl survey started in 1977 and expanded first in 1984 and later in 2003. Coverage has been inconsistent temporally and spatially from year to year. The gear has been consistent throughout the years (16 foot semi-balloon otter trawl). This data set was standardized using a delta GLM method, which is a two-stage approach that models presence and absence in the first stage and the abundance in the second. Several covariates were used in the analysis, including year, month, strata, salinity, temperature and depth. This approach seems appropriate, but no interaction terms were used in the standardization. They should be included.

Winter Dredge Survey

This is by far the best available set of information from the assessment. It covers the entire bay, at a moment the crabs become quiescent and are closely related or buried in the sediments. The survey has been running since 1991 and, on average, 1200 stations are visited each year. Each year trials are conducted to estimate vessel and year-specific catchability coefficients. These coefficients with area swept are used to come up with absolute abundance.

During the meetings the review panel members were suspicious about using this time series as an absolute index of abundance in the assessment. When this assumption was relaxed and q was estimated for this survey, the fit to different pieces of information improved substantially.

ToR D: Describe and quantify patterns in catch and effort by sector and region, including analyses that examine the impacts of reporting changes and trends in CPUE.

The assessment report considers the description/analysis of catch data of three different jurisdictions. Collecting information from different jurisdictional agencies, processing the data and interpreting the results was probably a major task in itself. In the assessment report only catch information was available for this analysis. Effort and CPUE data were excluded from the report. This is unfortunate, because it is always valuable to have fisheries indicators that directly relate to the fishery. It is important that the collection and analysis of such data be prioritized. The availability of such data becomes even more important if economic assessments are needed in the near future.

Fisheries in the Chesapeake Bay encompass a diversity of sectors, including recreational and commercial fisheries. Catch statistics come from 3 different jurisdictions (Virginia, Maryland and Potomac River fisheries commission), Virginia and Maryland being the most prominent ones. The most important concerns with respect to Chesapeake Bay blue crab catch data is misreporting of landings or missing entire sections of landings.

Changes in the reporting system and how this has been taken into account has been a point of controversy in previous assessments. The current assessment model does not

keep track of gear, fleet or jurisdiction. Therefore it requires only one overall catch time series. The problem of adjusting the reporting systems in the Virginia and Maryland jurisdictions remains and the current time series approach to deal with the 1993 discontinuity seems appropriate. Nevertheless the adjustments levels in the past catch records are substantial, so three additional procedures are recommended:

a) Do additional model runs with unadjusted catch levels to assess the sensitivity to this scenario.

b) Use unadjusted catch records, but use an additional intervention parameter in the model to deal with this discontinuity. This approach should pass some of the uncertainty around changes in reporting system into the assessment.

c) Look for independent data to verify the year and level of the scaling factor needed to correct the time series of catch.

The assessment model estimates observation and process error simultaneously. As part of model development it would be convenient to fix the catch related observation error by assigning specific coefficient of variation values to different parts of the catch time series, reflecting the relative uncertainty (confidence) around those estimates.

The lack of biological data (i.e. size and size-weight) in the historical catch records is also a weak point of fisheries related information of this stock. Catch in weight is converted to numbers, before being used in the assessment model, by using a constant scaling factor throughout the years, as opposed to each year's average weight. This point deserves some attention.

Other relevant current and historical factors affecting overall catch levels are:

a) Soft Peeler fishery data was not incorporated in the analysis.

b) Recreational catch underreporting, especially in some areas.

    - New sensitivity runs for different levels of recreational catch seem appropriate.

    - Develop a second phone survey.

c) Incidental mortality not being adequately addressed:

    - Mortality from the dredge winter fishery. Testimonial of some watermen before the fishery closure indicated a high Age-0 female incidental mortality. Some sensitivity accounting for some levels of additional mortality are recommended.

    - Summer pot fishery had up to 50-70% of incidental mortality.

    - In the trotline fishery.

ToR E: Develop and implement assessment models for the Chesapeake blue crab fisheries. In particular, models that permit estimates of the trends and status of the crab population and fisheries on a sex-specific basis should be evaluated.

The new assessment is an improvement with respect to its predecessors. This two-stage model keeps track of sex-specific abundance in the population and the catch, and the "age-structure" matches up with the available structured information from the surveys. This is a good example of a parsimonious model, which balances out complexity and simplicity. The renewal function is a modified Ricker model with both sexes affecting recruitment at larger population sizes and only females at low population levels. The rationale for the development of this recruitment function was discussed during the meeting and it seems to be valid, but in practical terms the recruitment function does not pass through the origin under some modeling conditions. In fact, as reported in the document yield is maximized when males are fished out of the populations. This is not a very precautionary model configuration. Some modifications in the model need to be introduced to avoid such non-precautionary behavior.

Under the current management scheme for the Chesapeake Bay blue crab fishery, the assessment model is only used for reference point calculation and not for stock status estimation. Models are run every five to eight years, and abundance is derived from an empirically based winter dredge survey (WDS) on a yearly basis. This procedure heavily relies on the assumption of WDS to be an absolute index of abundance. This issue deserves some additional investigation and to evaluate the option of running yearly assessments to yield model based stock status estimates.

The initial conditions of the assessment model were not documented in the assessment model. This needs to be corrected in the document. A direct inspection of the model equations in the ADMB code revealed that the initial condition assumptions are not fully consistent with the rest of the model's configuration. This issue will probably not be very influential in stock assessment results, given the short lifespan of this species, but it should be corrected for model consistency.

As pointed out previously, fecundity needs to be considered in the model for spawning stock calculation. Under the current model configuration fecundity is assumed to be constant across ages of terminal molted females. This will probably lead to an overestimation of spawning stock size, as fecundity should drop in female blue crab after the first reproductive season. The effect of this on reference point calculation is yet to be evaluated.

In the future, improvements to the stock assessment model should consider:

a) The use of size structure in the assessment. This can be done either by keeping the age structure of the model and using size data as an additional source of information (Fournier et al. 1998) or migrating into a fully size-structured model (Sullivan et al. 1990).

b) Include gear type through gear specific selectivity parameters. Under the current model configuration this could be done by re-parameterizing the partial recruitment parameter in the demographic equations by growth related and selectivity components. Under this new configuration the evaluation of management tactics and strategies on different gears types and its effect on the blue crab population dynamics will be readily available.

c) Some degree of spatial component to address different jurisdictions and something that takes into account key elements of the complex spatial dynamics of this species (i.e. upper portion of the Bay as mating grounds and lower portion of the Bay as a natural repository of mature females). The appropriate number of spatial components will emerge from a balance between important spatial dependent processes that need to be modeled and the amount of available data to parameterize these models.

Two other assessment models were presented in the stock assessment report, an update of the 2005 assessment and a biomass dynamic model. The different metrics of these models make them hard to compare with the new model.

ToR F: Examine density-dependent exploitation patterns derives from survey-based and model-based approaches.

The paper by Lipcius (Assessment working paper 2) presents convincing evidence of density-dependent exploitation operating on the Chesapeake Bay blue crab stock. The author used a set of linear and non-linear models to test for density-dependent harvest by using the winter dredge survey data. Therefore his analysis was entirely driven by survey-based data. No analysis based on the assessment's model-based abundance and exploitation rate estimates were presented.

The author should include a table with parameter estimates and standard error, as well as likelihood and AIC values. This will allow the reader to quantitatively judge the amount of evidence in favor of each model.

ToR G: Characterize scientific uncertainty with respect to assessment inputs and stock status.

The current assessment model is used for reference point calculation, as stock status is based on the winter dredge survey abundance estimates.

Uncertainty in a stock assessment model can be of various types, including for example model and parameter uncertainty, observation and process error (Francis and Shotton 1997). Model uncertainty was somehow covered in the current assessment, by using 3 different stock assessment models. The results were not fully compared, because they have different metrics.

Parameter uncertainty was addressed by a sensitivity analysis considering the following varying factors:

- Levels of natural mortality (M).
- Estimation of male natural mortality (M_{male}).
- Levels of recruitment parameter (r).
- Estimation of sex ratio.
- Estimation of recruitment parameter.

The results show low levels of variation in the point estimates of biological reference points (u_{MSY} varying from 0.3-0.36).

As part of model development, it seems appropriate to extend this effort by increasing the number of scenarios to include:

- Sensitivity to different model parameters.
- Assess the consistency of the results (stock status and reference points) by fitting the model to different pieces of information separately.
- Using different catch histories (or estimating intervention parameter).
- Estimation of either process or observation error in the model, at various levels. With the current configuration it is hard to infer the implications of different assumptions (observation/process error levels) in that regard.

During the assessment review, the review panel asked the stock assessment modelers to re-run the model under the assumption that the winter dredge survey was not an absolute index of abundance for Age-1+ crabs. The results are summarized in the following table:

Name	Mf	Mm	rf	sex r	neg LL	Nf2010	Frat 68-06	Frat2009	Nf(MSY)	Nf2009/Nf targ2009	uMSY2009	utarg2009
Base	0.9	0.90	0.6	0.52	248.4	181	0.97	1.39	158	0.84	0.34	0.25
low M	0.6	0.60	0.6	0.52	232.9	179	1.21	1.70	183	0.73	0.38	0.28
High M	1.2	1.20	0.6	0.52	246.2	182	0.84	1.25	145	0.92	0.30	0.22
low r	0.9	0.90	0.3	0.52	252.7	181	0.99	1.42	174	0.80	0.33	0.24
high r	0.9	0.90	0.9	0.52	247.1	180	0.97	1.37	153	0.86	0.34	0.26
est male M	0.9	1.39	0.6	0.52	86.7	197	0.73	1.72	211	0.68	0.35	0.26
est sex ratio	0.9	0.90	0.6	0.64	78.3	199	0.90	2.19	245	0.60	0.36	0.27
est rf	0.9	0.90	1.0	0.52	246.9	180	0.97	1.37	150	0.86	0.34	0.26
est WDS q	0.9	0.9	0.6	0.52	41.9	15.8	0.4	0.5	45.7	0.3	0.65	0.45
comb SR	0.9	0.9	0.6	0.52	248.5	180.3	0.9	1.4	129.3		0.34	0.25

These results are very different than any of the previous sensitivity tests. This is a main point of concern, because assuming that an index of abundance is absolute as opposed to relative is the exemption rather than the norm. The fit to the different pieces of information is much better under this scenario than the base case. It is highly recommended that the dredge survey assessment team discuss this issue with the stock assessment team to find plausible explanations. Under the *q different from 1* scenario the stock status time series index derived from the survey needs to be properly adjusted by the new catchability coefficient.

Uncertainty around model parameter estimates, model-based stock status and reference points were not provided in the document. Reporting partial likelihoods associated with each piece of information is also valuable.

Reporting the partial likelihoods is highly recommended in order to capture the uncertainty under different scenarios.

ToR H: Evaluate stock status with respect to reference points.

The assessment team presented a consistent way of calculating model-based biological reference points, and then a reasonable procedure for calculating target and limit reference points. This information was appropriately summarized in figure 1.

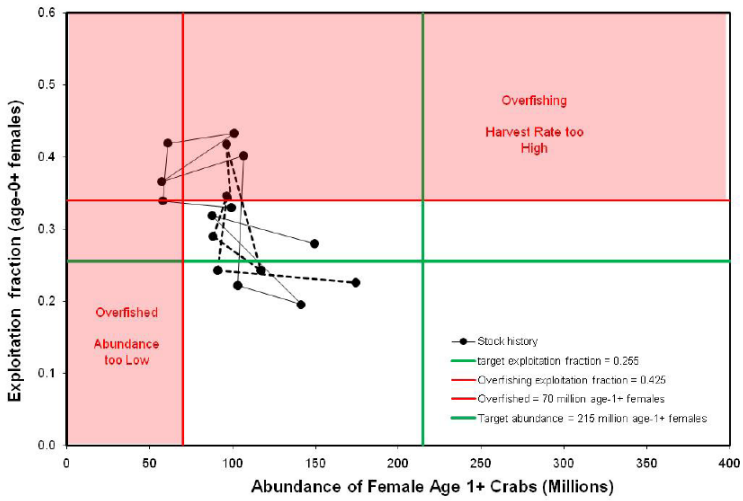

Figure 1: Control rule extracted from the 2010 Chesapeake Bay blue crab assessment (Figure 6.14 from the main assessment report).

Under the base case scenario, the stock in recent years is neither experiencing overfishing nor is it overfished.

The following recommendations emerged during the review to improve the representation and interpretation of the results:

- Include a year-caption associated to each stock status data point.
- A composite reference point figure, including: a) the 2005 method for standardizing the survey data and 2005 reference points, b) the 2010 standardized data with 2005 reference points and c) the 2010 standardized data with the 2010 reference points. This figure should be accompanied by an interpretation of the changes when moving from the 2005 to the 2010 assessment.

Preliminary extra model runs with estimated winter dredge survey catchability show different reference points and much higher exploitation rates throughout the years. This is presented in Figure 2. Under this scenario the stock has been experiencing

overfishing and has been overfished for a great portion of the time series. This point requires further discussion, as the model results are extremely sensitive to this scenario.

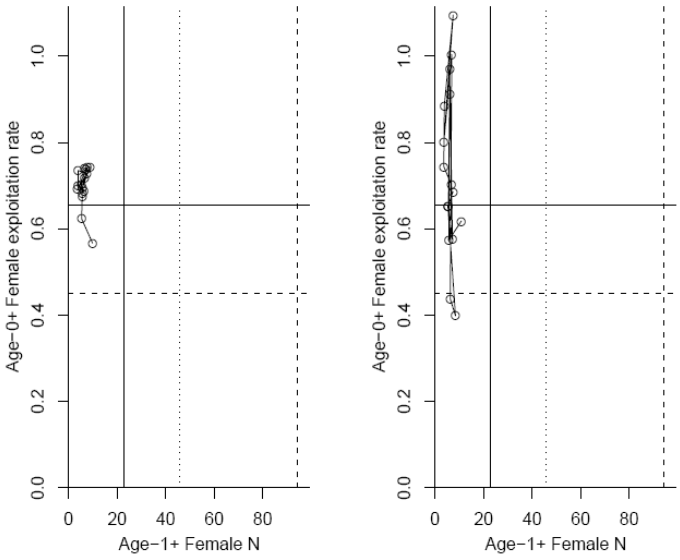

Figure 2: Recommended revised control rule under the estimated q for the winter dredge survey scenario. Figures were provided during the meeting from additional model runs. Vertical and horizontal lines represent limit an target reference points.

Acknowledgements

I would like to thank Derek Orner from the NOAA Chesapeake Bay Office for the coordination of the review process in Baltimore and for inviting us to the National Shellfisheries Association Symposium. He was very helpful and supportive. Tom Miller (University of Maryland) did a great job summarizing the assessment, Michael Wilberg was very responsive in rerunning the model and Romuald Lipcius gave great feedback on the winter dredge survey program. The entire assessment team was very open to questions raised by the review panel during the meeting. I would like to thank Manoj Shivlani (CIE), for contacting me and taking care of contractual and practical arrangements in an efficient manner.

Bibliography

Ernst, B., Orensanz, J. and Armstrong, D. 2005. Spatial dynamics of female snow crab (*Chionoecetes opilio*) in the Eastern Bering Sea. CJAFS 62:250-268.

Fournier, D., Hampton, J., Sibert, J. 1998. MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, *Thunnus alalunga*. Canadian Journal of Fisheries and Aquatic Sciences, 55:2105-2116.

Francis, R.I.C.C. and R. Shotton, R. 1997. "Risk" in fisheries management: a review. Can. J. Fish. Aquat. Sci. 54: 1699–1715.

Frusher SD & Hoenig JM (2001). Estimating natural and fishing mortality and tag reporting rate from up to three tagging events within a year in a multi-year tagging model. Canadian Journal of Fisheries and Aquatic Sciences. 58: 2490–2501.

Ju, S.J., D. H. Secor, and H. R. Harvey. 1999. Use of extractable lipofuscin for age determination of blue crab *Callinectes sapidus*. Marine Ecology Progress Series 185:171-179.

Lipcius, R. and Smith, A. Survival. 2011. Longevity and natural mortality of mature female blue crabs. Assessment Working Paper 1. Virginia Institute of Marine Science College of William & Mary Gloucester Point, VA 23062.

Lipcius, R. 2011. Density-dependent exploitation of blue crab in Chesapeake Bay. Assessment Working Paper 1. Virginia Institute of Marine Science College of William & Mary Gloucester Point, VA 23062.

Maunder, M. 2001. Integrated tagging and catch-at-age analysis (ITCAAN): model development and simulation testing. Spatial Processes and Management of Marine Populations 123 Alaska Sea Grant College Program. AK-SG-01-02.

Orensanz, J., Ernst, B., Armstrong, D. and Parma, A. 2005. Detecting early warnings of recruitment overfishing in male-only crab fisheries. In Fisheries Assessment and Management in Data-Limited Situations. Alaska Sea Grant College Program AKSG-05-02. University of Alaska, Fairbanks. Alaska.

Pascual, M. and Iribarne, O. 1993. How good are empirical predictions of natural mortality? Fisheries Research 16(1): 17-24.

Prager, M.H., McConaugha, J.R., Jones, C.M. & Geer, P.J., 1990. Fecundity of blue crab, *Callinectes sapidus*, in Chesapeake Bay: Biological, statistical and management considerations. Bulletin of Marine Science, 46, 170-179.

Puckett, B. J., D. H. Secor, and S. J. Ju. 2008. Validation and application of lipofuscin based age determination for Chesapeake Bay blue crabs (*Callinectes sapidus*). Transactions of the American Fisheries Society 137:1637-1649.

Smith, A. D., Sainsbury, K. J. and Stevens, R.A. 1999. Implementing effective fisheries-management systems – management strategy evaluation and the Australian partnership approach. ICES J. Mar. Sci. 56 (6): 967-979.

Sullivan, P., HAN-LIN Lai, Gallucci, V. 1990. A catch-at-length analysis that incorporates a stochastic model of growth. Canadian journal of fisheries and aquatic sciences 47(1):184-198.

Attachment A: Statement of Work for Dr. Billy Ernst

External Independent Peer Review by the Center for Independent Experts

Blue Crab Benchmark Stock Assessment - 2010

Scope of Work and CIE Process: The National Marine Fisheries Service's (NMFS) Office of Science and Technology coordinates and manages a contract providing external expertise through the Center for Independent Experts (CIE) to conduct independent peer reviews of NMFS scientific projects. The Statement of Work (SoW) described herein was established by the NMFS Project Contact and Contracting Officer's Technical Representative (COTR), and reviewed by CIE for compliance with their policy for providing independent expertise that can provide impartial and independent peer review without conflicts of interest. CIE reviewers are selected by the CIE Steering Committee and CIE Coordination Team to conduct the independent peer review of NMFS science in compliance the predetermined Terms of Reference (ToRs) of the peer review. Each CIE reviewer is contracted to deliver an independent peer review report to be approved by the CIE Steering Committee and the report is to be formatted with content requirements as specified in **Annex 1**. This SoW describes the work tasks and deliverables of the CIE reviewer for conducting an independent peer review of the following NMFS project. Further information on the CIE process can be obtained from www.ciereviews.org.

Project Description: The blue crab stock has been subject to Baywide stock assessments on two previous occasions. In the years between benchmark assessments, updates on the stock status are provided by the NOAA Chesapeake Bay Office's Chesapeake Bay Stock Assessment Committee. The most recent update concluded that the stock was not overfished and was not then experiencing overfishing. Since the 2005 assessment, the three management jurisdictions have implemented a range of regulatory changes aimed at attaining the target exploitation rate of 46% of the available stock. Thus, it is appropriate that another, Baywide benchmark assessment be conducted. The blue crab resource, specifically for soft and peeler crabs, in Chesapeake Bay has recently been declared a fisheries resource disaster by the Secretary Commerce. In 2009 and 2010, annual updates (not peer-reviewed) have shown slight improvements in the resource. Blue crab is the most important commercial fishery in Chesapeake Bay with annual Baywide landings recently as low as 50 million pounds – roughly 25 million pounds below the long-term average. 2010 predicted landings (if fished at the target exploitation level) could top 100 million pounds. This is obviously a large fluctuation in landings and thus value of the resource to the Bay community.

The first Baywide stock assessment was conducted using a length-based approach to estimate exploitation, and an unweighted average of the four principal fishery-independent surveys to determine abundance. Consequently biological reference points were crude.

In 2001, the technical subcommittee of the Bi-State Blue Crab Advisory Committee (BBCAC) developed a new management framework that relied on exploitation and biomass threshold and target reference points. Threshold reference points were proposed based on a maintaining 10% of the virgin spawning potential and on the lowest observed abundance in the surveys. A target exploitation rate that would lead to

an effective doubling of the spawning stock present in 2001 was also selected. The most recent Baywide benchmark assessment for blue crab in the Chesapeake Bay was conducted in 2005. This assessment critically evaluated and revised estimates of the natural mortality rate, the impact of reporting changes on landings estimates, and spawning potential ratio reference points. The 2005 assessment, using data through 2003, recommended adopting the exploitation fraction, defined as the proportion of crabs available at the beginning of the season that are subsequently harvested, in place of less intuitive measures (F) used in previous assessments. Estimates of exploitation fractions were calculated based on the Baywide winter dredge survey (WDS) and within a modified catch-survey analysis that permitted the use of multiple surveys. The approach used in the 2005 assessment was reviewed by a panel of international scientists with expertise in crustacean fisheries who found that it was a substantial improvement over previous assessments. However, the panel also identified issues to be addressed in future assessments. In particular, the panel recommended exploration of the impact of density-dependent processes in life history traits, improvements to the fishery-independent surveys, particularly with regard to catchability, the possibility of developing a sex-specific assessment model and reference points, and a fuller analysis of the impacts of uncertainty on all aspects of the assessment.

The 2010 assessment and targeted research program is a highly collaborative and integrated program to address specific concerns raised by the international review panel from 2005.
The assessment activities are divided into eight specific Terms of Reference (TOR) that were developed based on the review comments received from panel of experts convened to review the 2005 assessment, and from extensive discussion with managers from MDNR, the Potomac River Fisheries Commission and the Virginia Marine Resources Commission, the three relevant management jurisdictions.

NOAA Fisheries is playing a significant role in coordinating disaster assistance to Maryland and Virginia to ensure a sustainable blue crab fishery in Chesapeake Bay. This 2010 Benchmark assessment and research program represents a large investment by NOAA and the state management agencies and should be reviewed internationally.

The Terms of Reference (ToRs) of the peer review are attached in **Annex 2**. The tentative agenda of the panel review meeting is attached in **Annex 3**.

Requirements for CIE Reviewers: Three CIE reviewers shall conduct an impartial and independent peer review in accordance with the SoW and ToRs herein. CIE reviewers shall have working knowledge and recent experience in the application of stock assessment and crustacean fisheries. Each CIE reviewer's duties shall not exceed a maximum of 14 days to complete all work tasks of the peer review described herein.

Location of Peer Review: Each CIE reviewer shall conduct an independent peer review during the panel review meeting scheduled in Baltimore, Maryland during the tentative dates of 29-31 March 2011.

Statement of Tasks: Each CIE reviewers shall complete the following tasks in accordance with the SoW and Schedule of Milestones and Deliverables herein.

Prior to the Peer Review: Upon completion of the CIE reviewer selection by the CIE Steering Committee, the CIE shall provide the CIE reviewer information (full name, title, affiliation, country, address, email) to the COTR, who forwards this information to the NMFS Project Contact no later the date specified in the Schedule of Milestones and Deliverables. The CIE is responsible for providing the SoW and ToRs to the CIE reviewers. The NMFS Project Contact is responsible for providing the CIE reviewers with the background documents, reports, foreign national security clearance, and other information concerning pertinent meeting arrangements. The NMFS Project Contact is also responsible for providing the Chair a copy of the SoW in advance of the panel review meeting. Any changes to the SoW or ToRs must be made through the COTR prior to the commencement of the peer review.

Foreign National Security Clearance: When CIE reviewers participate during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for CIE reviewers who are non-US citizens. For this reason, the CIE reviewers shall provide requested information (e.g., first and last name, contact information, gender, birth date, passport number, country of passport, travel dates, country of citizenship, country of current residence, and home country) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: http://deemedexports.noaa.gov/sponsor.html).

Pre-review Background Documents: Two weeks before the peer review, the NMFS Project Contact will send (by electronic mail or make available at an FTP site) to the CIE reviewers the necessary background information and reports for the peer review. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the CIE Lead Coordinator on where to send documents. CIE reviewers are responsible only for the pre-review documents that are delivered to the reviewer in accordance to the SoW scheduled deadlines specified herein. The CIE reviewers shall read all documents in preparation for the peer review.

Panel Review Meeting: Each CIE reviewer shall conduct the independent peer review in accordance with the SoW and ToRs, and shall not serve in any other role unless specified herein. **Modifications to the SoW and ToRs can not be made during the peer review, and any SoW or ToRs modifications prior to the peer review shall be approved by the COTR and CIE Lead Coordinator.** Each CIE reviewer shall actively participate in a professional and respectful manner as a member of the meeting review panel, and their peer review tasks shall be focused on the ToRs as specified herein. The NMFS Project Contact is responsible for any facility arrangements (e.g., conference room for panel review meetings or teleconference arrangements). The NMFS Project Contact is responsible for ensuring that the Chair understands the contractual role of the CIE reviewers as specified herein. The CIE Lead Coordinator can contact the Project Contact to confirm any peer review arrangements, including the meeting facility arrangements.

Contract Deliverables - Independent CIE Peer Review Reports: Each CIE reviewer shall complete an independent peer review report in accordance with the SoW. Each CIE reviewer shall complete the independent peer review according to required format

and content as described in Annex 1. Each CIE reviewer shall complete the independent peer review addressing each ToR as described in Annex 2.

Other Tasks – Contribution to Summary Report: Each CIE reviewer may assist the Chair of the panel review meeting with contributions to the Summary Report, based on the terms of reference of the review. Each CIE reviewer is not required to reach a consensus, and should provide a brief summary of the reviewer's views on the summary of findings and conclusions reached by the review panel in accordance with the ToRs.

Specific Tasks for CIE Reviewers: The following chronological list of tasks shall be completed by each CIE reviewer in a timely manner as specified in the **Schedule of Milestones and Deliverables**.

1) Conduct necessary pre-review preparations, including the review of background material and reports provided by the NMFS Project Contact in advance of the peer review.
2) Participate during the panel review meeting in the Baltimore, Maryland during the tentative dates of 29-31 March 2011, and conduct an independent peer review in accordance with the ToRs (**Annex 2**).
3) No later than 14 April 2011, submit an independent peer review report addressed to the "Center for Independent Experts," and sent to Mr. Manoj Shivlani, CIE Lead Coordinator, via email to shivlanim@bellsouth.net, and Dr. David Sampson, CIE Regional Coordinator, via email to david.sampson@oregonstate.edu. Each CIE report shall be written using the format and content requirements specified in Annex 1, and address each ToR in **Annex 2**.

Schedule of Milestones and Deliverables: CIE shall complete the tasks and deliverables described in this SoW in accordance with the following schedule.

22 February 2011	CIE sends reviewer contact information to the COTR, who then sends this to the NMFS Project Contact
15 March 2011	NMFS Project Contact sends the CIE Reviewers the pre-review documents
29-31 March 2011	Each reviewer participates and conducts an independent peer review during the panel review meeting
14 April 2011	CIE reviewers submit draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator
28 April 2011	CIE submits CIE independent peer review reports to the COTR
5 May 2011	The COTR distributes the final CIE reports to the NMFS project contact

Modifications to the Statement of Work: Requests to modify this SoW must be approved by the Contracting Officer at least 15 working days prior to making any permanent substitutions. The Contracting Officer will notify the COTR within 10 working days after receipt of all required information of the decision on substitutions. The COTR can approve changes to the milestone dates, list of pre-review documents, and ToRs within the SoW as long as the role and ability of the CIE reviewers to complete the deliverable in accordance with the SoW is not adversely impacted. The SoW and ToRs shall not be changed once the peer review has begun.

Acceptance of Deliverables: Upon review and acceptance of the CIE independent peer review reports by the CIE Lead Coordinator, Regional Coordinator, and Steering Committee, these reports shall be sent to the COTR for final approval as contract deliverables based on compliance with the SoW and ToRs. As specified in the Schedule of Milestones and Deliverables, the CIE shall send via e-mail the contract deliverables (CIE independent peer review reports) to the COTR (William Michaels, via William.Michaels@noaa.gov).

Applicable Performance Standards: The contract is successfully completed when the COTR provides final approval of the contract deliverables. The acceptance of the contract deliverables shall be based on three performance standards:
(1) each CIE report shall completed with the format and content in accordance with **Annex 1**,
(2) each CIE report shall address each ToR as specified in **Annex 2**,
(3) the CIE reports shall be delivered in a timely manner as specified in the schedule of milestones and deliverables.

Distribution of Approved Deliverables: Upon acceptance by the COTR, the CIE Lead Coordinator shall send via e-mail the final CIE reports in *.PDF format to the COTR. The COTR will distribute the CIE reports to the NMFS Project Contact and Center Director.

Support Personnel:

William Michaels, Program Manager, COTR
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
William.Michaels@noaa.gov Phone: 301-713-2363 ext 136

Manoj Shivlani, CIE Lead Coordinator
Northern Taiga Ventures, Inc.
10600 SW 131st Court, Miami, FL 33186
shivlanim@bellsouth.net Phone: 305-383-4229

Roger W. Peretti, Executive Vice President
Northern Taiga Ventures, Inc. (NTVI)
22375 Broderick Drive, Suite 215, Sterling, VA 20166
RPerretti@ntvifederal.com Phone: 571-223-7717

Key Personnel:

NMFS Project Contact:
Derek M. Orner
NOAA Chesapeake Bay Office, 410 Severn Avenue, Suite 107A, Annapolis, MD 21403
Derek.orner@noaa.gov Office: (410) 267-5676 Cell: (410) 570-2268

Annex 1: Format and Contents of CIE Independent Peer Review Report

1. The CIE independent report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether the science reviewed is the best scientific information available.
2. The main body of the reviewer report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs.
   a. Reviewers should describe in their own words the review activities completed during the panel review meeting, including providing a brief summary of findings, of the science, conclusions, and recommendations.
   b. Reviewers should discuss their independent views on each ToR even if these were consistent with those of other panelists, and especially where there were divergent views.
   c. Reviewers should elaborate on any points raised in the Summary Report that they feel might require further clarification.
   d. Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.
   e. The CIE independent report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed, regardless of whether or not they read the summary report. The CIE independent report shall be an independent peer review of each ToRs, and shall not simply repeat the contents of the summary report.
3. The reviewer report shall include the following appendices:
   Appendix 1: Bibliography of materials provided for review
   Appendix 2: A copy of the CIE Statement of Work
   Appendix 3: Panel Membership or other pertinent information from the panel review meeting.

Annex 2: Terms of Reference for the Peer Review

Blue Crab Benchmark Stock Assessment - 2010

The stock assessment review has the following eight specific terms of reference:

a) Critically assess and where necessary revise the life history and vital rates of blue crab in the Chesapeake Bay that are relevant to an assessment of the stock.

b) Evaluate and recommend biological reference points for the Chesapeake Bay blue crab population. The potential for implementing sex-specific reference points should be evaluated.

c) Describe and quantify patterns in fishery-independent surveys. Analyses should include an evaluation of the impacts of environmental and abiotic factors on survey catches, to maximize the information content of resultant survey time series.

d) Describe and quantify patterns in catch, effort and survey-based estimates of exploitation by sector and region, including analyses that examine the impacts of reporting changes and trends in CPUE.

e) Develop and implement assessment models for the Chesapeake blue crab fisheries. In particular, models that permit estimates of the trends and status of the crab population and fisheries on a sex-specific basis should be evaluated.

f) Examine density-dependent exploitation patterns derived from survey-based and model-based approaches.

g) Characterize scientific uncertainty with respect to assessment inputs and stock status.

h) Evaluate stock status with respect to reference points.

Annex 3: Tentative Agenda
2010
Blue Crab Stock Assessment Review
Sheraton Baltimore City Center Hotel
101 West Fayette St., Baltimore, MD
March 29-31, 2011

March 29, 2011

12:30 Welcome & Introductions
Orner
- Stock Assessment Committee
- Review Panel

12:45 Presentation of the 2010 Blue Crab Stock Assessment
Miller

4:00 General / Open Question Period
Orner
- Public Comment
- Review Panel

5:30 Adjourn

March 30, 2011

8:30 Term of Reference Review and Discussion

I. Critically assess and where necessary revise the life history and vital rates of blue crab in the Chesapeake Bay that are relevant to an assessment of the stock.

II. Evaluate and recommend biological reference points for the Chesapeake Bay blue crab population. The potential for implementing sex-specific reference points should be evaluated.

III. Describe and quantify patterns in fishery-independent surveys. Analyses should include evaluation of the impacts of environmental and abiotic factors on survey catches, to maximize the information content of resultant survey time-series.

IV. Describe and quantify patterns in catch and effort by sector and region, including analyses that examine the impacts of reporting changes and trends in CPUE.

12:30 Lunch

1:30 Term of Reference Review and Discussion (continued)

V. Develop and implement assessment models for the Chesapeake blue crab fisheries. In particular, models that permit estimates of the trends and status of the crab population and fisheries on a sex-specific basis should be evaluated.

VI. Examine density-dependent exploitation patterns derives from survey-based and model-based approaches.

VII. Characterize scientific uncertainty with respect to assessment inputs and stock status.

VIII. Evaluate stock status with respect to reference points.

5:15 Adjourn

March 31, 2011

9:00	Review Session *[closed-door]* - Review Panel to discuss assessment methodologies and develop individual opinions. - Initiate development of summary documents
12:00	Lunch
1:15	Review Session (continued)
4:30	Adjourn

Stock Assessment of Blue Crab in Chesapeake Bay 2011: Response to Reviewers

The 2011 Stock Assessment of Blue Crab in Chesapeake Bay was reviewed by a CIE panel on March 29-30, 2011 in Baltimore, MD. The review meeting involved a presentation by the assessment team to the panel on the first morning, followed by a day and half of detailed questions and answers involving the both the review team and the panel. In many cases the assessment team responded directly to questions raised by the panel by conducting additional simulation runs or analyses.

Each of the reviewers fully supported the approach taken in the assessment. They recognize that the assessment provides the "best available" science to support management of the blue crab fishery. There are no fatal flaws identified by any of the reviewers in their comments. We note that because of the nature of this specific review panel, there was no review panel chair and thus there is not a summary report. Yet, Dr. Addison sums up the reviewers comments when he writes,

"I would fully support the approach taken in the assessment. The derivation of the reference point calculations considers the sex-specific nature of the stock-recruitment relationship and is implemented correctly in the model. The exploitation and abundance reference points are meaningful indices which are understood readily by all stakeholders. In particular, the exploitation fraction (U) is a much simpler concept to understand than reference points based on fishing mortality (F). This is particularly important if management action is required when thresholds are exceeded or if targets are reached and agreement is needed on how to achieve the management objectives. I support the use of the US Federal definitions for the limit reference points and the use of the New England and Mid-Atlantic Fishery Management Councils' system to define target reference points. These reference points may seem a little arbitrary, but by their very nature, most reference points have a degree of subjectivity, and the authors have sensibly chosen to define their reference points along the lines of widely-agreed approaches."

However, each of reviewers does make important recommendations for immediate changes to the assessment document and recommendations for research that should be undertaken to improve future assessments. Here we respond to the specific points identified by the each individual review panel member in their CIE review. We address reviewer comments alphabetically in the order raised by each reviewer.

Julian Addison

Dr. Addison provided a 31-page review of the assessment. His recommendations are outlined in the review narrative and identified on a point by point basis in a list of 30 recommendations. There are no recommendations in the narrative that are not identified in the list of 30 recommendations.

A1. *Annual variations in growth rate in relation to the allocation of crabs from the surveys to age-0 and age-1+ age classes should be investigated to ensure that assumptions underlying the population model are correct.*

The review team recognizes that the allocation of crabs to either the age-0 or age-1+ category based on size introduces questions regarding the reliability of the size-age key. However, this issue affects the MD and VIMS surveys more than it does the winter dredge survey – for which the bimodal size distribution is remarkably constant from year to year. We do not believe this is an issue that affects the reliability of the current assessment.

The assessment team recommends that new research is conducted to explore the dynamics of blue crab size structure inter-annually as it relates to environmental parameters and population size.

A2. Sensitivity analysis of the assessment model outputs to the partial recruitment curve generated from the Puckett et al. study should be carried out.

The partial recruitment into the fishery likely varies inter-annually as a function of the growth of blue crabs during their first full summer when individual crabs grow rapidly from a carapace width of < 60 mm on emergence in April to attaining maturity by the autumn. In the base model we assumed that the partial recruitment of these age-1 crabs was 0.6. We addressed concerns raised by this reviewer by conducting sensitivity runs that are fully documented in the assessment. The three sensitivity runs include two runs with fixed partial recruitments of 0.3 and 0.9 which together with the base case fully bracket the range suggested in the Puckett et al. paper. We conducted an additional run in which we allowed the partial recruitment to be estimated. This simulation still used a constant partial recruitment value – but the specific value was estimated by the model. Model results from this run suggest the model performed best with a partial recruitment close to 1. We caution that this should not be taken to indicate that the true partial recruitment =1. We maintain that the base case assessment model with the partial recruitment =0.6 still represents the best available information for the true value of the partial recruitment.

The assessment team recommends that additional research to evaluate the utility of including a variable partial recruitment parameter be conducted.

A3. The population model should be modified to take into account the terminal moult in female crabs replacing the current assumption that all age-1+females contribute to the spawning stock.

The assessment team is not fully certain of the intent of this comment, but we believe that it relates to the fact that there is no age-specific decline in fecundity in the

assessment model. As currently implemented, the model assumes that were a female to survive to an age of 6 or 7, that female would have exactly the same fecundity as she experienced with her first brood. Empirical evidence from experience of raising crabs in the NOAA-funded crab hatchery in Baltimore and from field studies in other states suggest this is not the case.

The assessment team recognizes this deficiency and suggests that it could only be addressed by a division of the age-1+ category in the model into two separate stages – first time spawners and experienced spawners. Although this would indeed provide additional resolution of age-specific fecundity, it would do so at a cost of compromising our ability to identify age categories in the survey data. Given the high mortality rate of females, however, a very small proportion of females is expected to survive long enough to experience reduced fecundity.

The assessment team believes that a more resolved stage- or age-structured model should be a goal for a future assessment.

A4. The potential for sperm limitation in blue crabs should be revisited particularly in the light of recent new management measures to protect the female stock which have caused an increase in the male exploitation rates.

Sperm limitation is known to occur in several crustacean species and has been explored experimentally and empirically for blue crab in Chesapeake Bay. It is clear from these studies that there is the potential for sperm limitation for occur and the presence of sperm limitation should be a concern for managers responsible for the blue crab fishery. However, evidence suggests that sperm limitation is currently not limiting population productivity in the Chesapeake Bay system. Provided that managers follow the guidelines in the assessment of maintaining the sex ratio in the population within observed bounds, concerns over sperm limitation should not restrict the freedom of managers to act.

The assessment team recommends that resources be invested to understand the scope for sperm limitation in blue crab in the Chesapeake Bay. Research could take advantage for example of the natural trend in sex ratios up and down the bay to assess the impact of a reduction in male abundance on female reproductive success.

A5. Use of the Brownie model with a different parameterisation that includes M and F, or M and catchability and effort to analyse tag-recapture returns would be informative.

The assessment team recognizes the important role that the tag-recapture data can play in informing future assessments. Such data could provide a time series of estimates of M to which a future model could be fit, or more simply could inform and improve estimates of M. The assessment team fully supports this recommendation.

A6. Natural mortality rate (M) can vary annually, seasonally and with crab size. It is recommended that an analysis is carried out to assess the sensitivity of the model's output to variable M.

It is highly likely that the assessment will be sensitive to the value of M used or to the existence of interannual variability in M. However, we currently have few data that inform us as to the magnitude and pattern of that variability. Accordingly the assessment team recommends that research to address this recommendation be tied directly to the previous recommendation to document patterns in the magnitude and variability in M.

A7. A spatial component should be added to the population model or at the very least some spatial analysis of the survey data should be undertaken.

The assessment team considered development of a spatially-explicit model early on in the assessment process, but rejected it as a viable alternative because we have insufficient knowledge regarding the movement of blue crab. Lacking such information, inclusion of a spatial component risked becoming a model fitting exercise, allowing the model much more freedom to fit observed pattern, but unconstrained by any information that would allow us to evaluate the reliability of the movement patterns implied from fits to a spatially-explicit model.

The assessment team continues to believe that a spatially-explicit model should be a goal for future assessments. Spatially-explicit data are available from fishery-independent surveys, and increasingly from fishery-dependent sources. We recommend that simulation modeling be undertaken to explore trade-offs in changing the spatial and temporal resolution of the existing model.

A8. Recent increases in male exploitation rate and the increased potential for sperm limitation, suggest that male-based reference points might be appropriate, for example, in relation to abundance of age-1+ male crabs.

No explicit male-specific reference points were brought forward in the model. However, male-specific reference points are implicit in the model through our strong caution that the ratio of sex-specific exploitation rates be maintained within the envelope of values that have been observed empirically. Thus, in setting a female-specific reference point and in maintaining observed ratios of sex-specific exploitation rates, managers are implicitly managing to a male-specific reference point.

The assessment team notes that male-specific reference points could be formally developed if so desired. But in keeping with points made by this same reviewer in his full narrative we believe that such reference points should come out of discussions between managers and stakeholders rather than from the assessment process alone.

A9. When standardising survey indices using the delta generalised linear model (deltaGLM) method, interaction terms in the second stage of the delta GLM should be considered, especially strata and year, temperature and year for the dredge survey.

The assessment recognizes that an interaction term would be appropriate in the standardization approach. The Delta-GLM package that is widely used within NMFS for stock assessments only permits first order effects to be included in the model. This does not excuse the oversight.

This was the first time that any of the fishery-independent indices have been standardized for a blue crab assessment. We believe that the standardization effort did improve the information provided to the assessment model. The standardized indices appear to describe the data well and we do not believe that inclusion of the interaction term would induce a substantial change in our conclusions. However, we do recognize that the inclusion of interactions may have improved the information content of the surveys even more. Therefore, the assessment team fully endorses the continued use of standardization analyses for future assessments. In addition, the assessment team recommends that future assessments should continue to include interaction terms in any effort to standardize fishery-independent surveys.

A10. The winter dredge survey (WDS) is an essential component of the assessment and the evaluation of stock status relative to the reference points and its continuation is strongly recommended.

The assessment team fully supports this recommendation. The winter dredge survey is the most accurate and comprehensive index we have available for this population and it plays an essential role in managing the population in between benchmark assessments.

A11. The Virginia (VIMS) trawl survey should be standardized, and better use should be made of the additional four strata.

The VIMS trawl survey is the longest single fishery-independent survey available and as such provides invaluable information to managers and assessment scientists. For this assessment, we did not make as full use of this index as perhaps we should have done – and we accept the criticism implied in this comment.

We note that the VIMS trawl survey is longer and its design has evolved and is accordingly more complex than the other surveys. The assessment team recommends that a high priority future research objective should be the comprehensive analysis of the full VIMS trawl blue crab database. To our knowledge this has not been done since Chris Bonzek (VIMS) re-evaluated the size-age class categorization as a part of the Rugolo et al. assessment. The application of modern standardization approaches to these data, and the use of these data to re-evaluate the size-age convention (points A1 and A2 above) would be of high utility to future modeling efforts.

A12. WDS survey results for Virginia and Maryland should be evaluated separately to investigate whether they correlate with trawl survey indices in the two states.

Such analyses have been done informally, but were not presented in the assessment. In general, the age-1+ abundances from the WDS correlate well with VIMS trawl survey indices and less well for the MD survey. Age-0 abundances from the WDS do not correlate well with recruitment indices from either state – further reinforcing the conclusion in the assessment regarding the low information content of these surveys for recruits. Additionally analyses have also correlated the mature female abundance from the winter dredge survey with the abundance of mature females on the spawning grounds the following spring and summer from the VIMS trawl survey. There are discrepancies that are apparent in this comparison that suggest processes may be occurring between the time when the females are surveyed in the winter dredge survey in the winter and when the females move onto the spawning grounds later in the year.

The assessment team does not believe there is much to be gained from the direct comparison of WDS and state surveys as recommended by the reviewer. However, the assessment team does suggest that continued evaluation of the relationship between mature female abundance in the dredge survey and in the VIMS trawl survey is warranted.

A13. Other survey time series, such as the Chesapeake Bay Multispecies Monitoring and Assessment Program, should potentially be included in the assessment.

The exclusion of the CHESMAPP survey data by the assessment team was a regrettable error. These data should have been evaluated. We chose not to evaluate them because we believed, *a priori*, that the time series was too short. However, this is a weak excuse for the lack of inclusion of these data and they should have been evaluated.

Since the assessment meeting we have obtained, courtesy of Dr. Rob Latour (VIMS), the full CHESMAPP blue crab data set for use in a separate production modeling exercise for blue crab (Colton, A. R. 2011. An Evaluation of the Synchronization in the Dynamics of Blue Crab Populations in the Western Atlantic. MS Thesis. University of Maryland College Park, College Park, MD.). Colton used a similar assessment model to that used here but included four fishery-independent indices for the Chesapeake (WDS, VIMS, MD Trawl and CHESMAPP). Her findings suggest that the utility of the CHESMAPP survey data for blue crab assessments is currently low, but we expect it will increase.

The assessment team recommends that any future assessment of blue crab explore and to the extent possible include the blue crab data from the CHESMAPP survey.

A14. Consideration should be given to the development of an early life history phase recruitment index, e.g. larvae or megalopae abundance.

The assessment team recognizes that only the WDS appears to provide reliable information on the abundance of recruits and thus, a survey specifically designed to sample blue crab megalopae as they return to the bay would be useful.

The assessment team notes that under one aspect of the funding obtained for the assessment, new work is being undertaken to improve the reliability of the WDS as an index of abundance. In this research, the WDS is being expanded into shallow water using a smaller dredge. Options of the design of such a survey, and preliminary estimates of the catchability of the gear are being considered now.

A15. The scale of changes to the time series of catch data following re-calibration to take into account changes in reporting procedures should be independently verified, even if indirectly and/or anecdotally.

The assessment team appreciates the spirit in which this comment is made, but believes that such a direct assessment would be fraught with difficulties. In considering review comments provided by all three reviewers, we believe a better approach would be to have used the time series analyses only to identify whether reporting changes were necessary, but to use the assessment model itself to estimate the magnitude of the reporting change by simply adding a parameter to the model itself in which the magnitude of the landings is scaled by a common factor before and after a prospective reporting change.

We propose that future assessments should implement the catch correction within the assessment model rather than external to it. We believe this would be an improvement because the time series approach used here was in itself a form of simple stock assessment – in that it assumed something about the relationship between catch and abundance. Using the output from such a model in a second assessment model does not permit for the error structure in the time series model to be brought forward into the assessment model

A16. The assessment team should gain access to, and analyse, the time series of catch data from the Potomac Fisheries Commission.

The assessment used the Potomac River data as the data from the Potomac is partitioned out to the two states. However, the reviewer is correct to the extent that the spatial and temporal resolution of the catch data from the PRFC is much greater than in either of the states. To the extent we did not do this, this was an opportunity lost. In particular, because the PRFC maintains good catch and effort data, we did indeed loose the opportunity of developing a commercial CPUE index that could have been used as an additional input to the model.

The assessment team recommends that future assessments explore data from the Potomac River to ensure catch and effort data available from this system are fully utilized.

A17. Accurate and complete catch and effort data should be collected from the fishery so that a consistent standardised catch rate dataset can be used in the assessment. The potential existence of personal diaries of daily catch and effort data should be investigated.

There have been considerable improvements to the quality of catch and effort data since the first assessment by Rugolo et al. in 1997. One of the ToRs of the project that funded the assessment was to analyze the existing effort data to determine if there was sufficient information in the data to form the basis of a CPUE index. We failed to complete this ToR.

18. A baywide survey of recreational catches should be undertaken immediately and at regular intervals to quantify the overall landings from the recreational sector.

The assessment team agrees that surveys of recreational removals should be a high research priority. We believe that such data would greatly improve the data available to managers and stakeholders to make informed decisions.

19. Some quantitative assessment of under-reporting rates should be obtained.

The extent of under-reporting is not quantified. Apocryphal tales suggests the magnitude of the problem may be substantial in some areas and at some times. The assessment team agrees that this should be a high priority research item.

20. Incidental mortality and discard mortality rates should be ascertained for the pot, trot line and dredge fisheries and survival rates for the peeler/soft crabs held in tanks should be qunatified.

Although poorly quantified for all sectors, the assessment team does not believe that the discard mortality rate in this fishery is substantial. However, we do support improvements to our understanding of the dynamic of the soft and peeler fishery.

A21. Detailed investigation is required as to why the model provides such a poor fit to the sex-specific population sizes.

We held long discussions with the review panel over the tension in the model introduced by the empirical patterns in the sex ratios in the surveys and the catches. We have tried numerous model runs with different parameterizations to try to account for this pattern in model results. The bias could be resolved by setting substantially natural mortality rates for males and females, or by establishing a sex-ratio at recruitment far different to those observed empirically. Both runs were not considered realistic. Overall, we developed no deep insight into either what are the biases inherent

in the data, or whether there is an important interaction among processes that occurs in the field but that is not fully captured in the model. We have no reason to believe that there is any coding error in the model that causes this pattern. Currently, we believe the discrepancy may result from patterns introduced into the data because of the catchability corrections for the winter dredge survey. Accordingly we believe this is a data issue more than it is a model issue. The assessment team strongly supports the research recommendation proposed that a detailed investigation be undertaken to understand the sources of the discrepancy.

A22. A conceptual model providing a timeline of life history events such as spawning and recruitment should be incorporated.

It was clear from the beginnings of our discussions with the assessment review panel that we had not done an adequate job of explaining the timing of key life history events or of key fishery activities. For the assessment presentation we developed a conceptual diagram of the model which helped greatly in resolving some of these issues. This is now included in the assessment document as Figure 6.3.

A23. A number of minor changes are required to the formulation of the model. These include defining the initial conditions of the model, decrementing M by time and F by k (kappa) in equations 6.7 and 6.11, incorporating a bias correction and incorporating a compensatory function into the stock-recruitment relationship.

The review panel and assessment teams had numerous exchanges during the Q & A period related to these points. The assessment team has formulated the two equations in question under the assumption that the temporal pattern of fishing mortality and natural mortality are identical and constant. If we accept this assumption, then the changes are not required.

The changes recommended by the review panel imply that the temporal trends in natural and fishing mortality are different. The assessment team accepts that this may be the case. But we would argue we do not have an *a priori* expectation for how M and F differ that would be required to follow the review panel's advice. We recommend viewing this question as a research recommendation rather than a short term change needed for this assessment.

The assessment team is not quite certain over the final two suggestions of the bias correction and the compensatory function. We believe the reviewer is misguided in his suggestion of the need for a bias correction in estimates. This is normally done when estimates from the assessment model are used in a separate framework to estimate reference points. In this case our reference points are estimated within the assessment model itself and thus we do not believe bias correction is necessary (i.e., the bias correction to calculate the mean of the alpha parameter of the stock-recruitment relationship from the median is done in the reference point calculations).

We believe the request for a compensation term in the stock recruitment relationship relates to the observation that the current stock recruitment model predicts maximum stock productivity at a zero male abundance level – something that is clearly biologically absurd. We have implemented this in additional sensitivity runs included in the assessment document and Assessment Working Paper III. We note that the uMSY reference point from this combined sex stock recruitment model is identical to that developed in the sex-specific base model. We further note that the non sex-specific stock recruitment model recommended by the review panel is not free of absurd implications in the face of extreme sex ratios. Specifically, because the sex composition of the spawning stock is not specified in the non sex-specific stock recruitment model, the implicit implication that average recruitment will be observed from stocks comprised of a normal sex ratio, an all female population and an **ALL** male population.

The assessment team continues to believe that the sex-specific stock recruitment model used in the assessment is more appropriate for blue crab than the non sex-specific model. To address the review panel's concerns we are currently working on a modified version of the sex-specific model that imposes a recruitment penalty on increasing female dominated populations to reflect sperm limitation. However this refinement of the model is not yet fully developed.

A24. The chosen values for the overfished abundance threshold and the target abundance reference point should be re-considered.

It is not clear to the assessment team exactly what the review is asking in this comment. We did have extensive discussions with the review team over assessment philosophy and whether it was the responsibility of the assessment team to offer managers with a range of options for reference points or a single value. Following federal guidelines, the assessment team believes that it should provide strong guidance with regard to limit reference points – albeit recognizing that even here there are no absolute standards and that any recommendation reflects the assessment history and practice for individual stocks. However, we recognized more flexibility with regard to target reference points. Paraphrasing John Hoeing (VIMS), "Science tells you want you can't have, not what you want." In this light, threshold reference points provide scientifically-based guidance as to limits for abundance and exploitation. The target reference point is simply one suggestion, following federal guidelines of what society might "want," but it is not the sole possible value.

The assessment team is happy to work with managers and stakeholders to explore a wider range of target reference points if so desired.

A25. Consideration should be given to incorporating more biological realism and structure into the model including a size-based approach, shorter time steps, and a spatial component.

The assessment team fully supports this recommendation for any future assessments. However, we would argue that we have taken a substantial step in developing sex-specific reference points in this assessment. We believe it is currently premature to go to a fully spatially resolved, high temporal resolution model.

A26. A more systematic sensitivity analysis should be undertaken to understand how robust the model's output is to the various input parameters.

Although the additional simulation runs required to address this review comment were not in the initial draft of the assessment made available to the review panel, such simulations were conducted during the review meeting and made available to the review panel. The results of these simulations are integrated into Table 6.2 and are available in Assessment Working Paper 3.

A27. The model should be run with and without the modified time series of catches, when observation error and process error are not estimated simultaneously, and with alternative configurations of the sex-specific stock recruitment relationship.

We have addressed aspects of this comment above in A15. We believe a better way in the future to assess the likely impact of the reporting change is to do so internally to the model. However, we do not believe it would be appropriate to do so here because it would substantially change the reviewed document and assessment.

We do now present an alternative run of the model with a traditional, non-sex specific stock recruitment model. We continue to argue that although traditional, the non-sex specific stock recruitment model is not free of assumptions about stock productivity at extreme sex ratios. The sex-specific model makes the claim that reproduction can occur in the absence of males – whereas the traditional model makes the claim that it can occur in the absence of females or males.

A28. One of the key assumptions of the model is that the abundance of age-1+ crabs from the WDS is an absolute estimate of abundance. As the assessment model is not robust to this assumption, the panel recommended that a sensitivity test is run that uses the raw winter dredge survey indices and that the derivation of the catchability of the winter dredge survey index of abundance should be investigated and possible sources of this difference (both assessment and survey) should be investigated.

We had extensive discussions with the review panel over the claim that the WDS is an index of absolute abundance. The review panel is correct in noting that surveys that are indices of absolute abundance are rare. Although the panel is certainly correct that such surveys are rare, we do feel that a good case can be made that the survey catchability based on the number of stations and coverage for age-1+ crabs is near unity. A simulation model run in which the catchability for the winter dredge survey was estimated (Table 6.2 – Est WDSq simulation) resulted in estimates of uMSY

exploitation of almost twice that of the other model runs. We believe adopting this as the base run to be a non-precautionary approach.

However, the assessment team fully supports additional studies to characterize the catchability of age-1+ adult crabs in the winter dredge survey. An additional benefit that might accrue from such research would be insights into the sex-ratio at recruitment knowledge of which might help us explain the failure to achieve sex-specific sex ratios in the data. We note that one aspect of the research funded alongside the assessment was to do just that.

A29. Additional interpretation should be provided about the two changes that occur between the calculation of the 2005 reference points and those in the 2010 assessment.

We believe that the reviewer is requesting that we run the 2005 assessment model with standardized indices and the 2010 assessment with the raw indices (as used in the 2005 assessment model) to evaluate whether the change in the assessment model or the change in indices is most responsible for the change in reference points. We believe such simulations are not necessary. Given the close agreement between the simple survey means and the standardized index values for most age classes and surveys, it is clear to the assessment team that the change in the assessment model is most responsible for the change in the reference points.

A30. Harvest control rules should be developed to agree management actions if reference points are reached or exceeded.

We agree with the review panel that development of a management control rule would be a helpful addition to the framework for managing blue crab in the Chesapeake.

Dr. Cathy Dichmont

Dr. Dichmont provided a detailed review of the assessment. In her review, she identified two categories of recommendations: future work and recommendations. It is clear that the two recommendations in the future work section relates to medium- to long-term investments that may provide additional flexibility to assessment analysts working on blue crab in the future. However, while many of the forty recommendations relate to specific actions that she recommends are addressed in the current assessment, the assessment team believes that some of these also are best addressed as targeted research initiatives to be conducted over the next five years to prepare the ground for the next assessment. In addition there is considerable overlap between comments from Drs. Dichmont and Addison. Where appropriate we refer the reader back to comments made in response to Dr. Addison's points.

We comment on each below

Future work beyond recommendations

1. *This assessment would benefit from being size---based with a shorter time step – this would internalize the growth rate assumptions, allow in---season changes to be accommodated and also incorporate different gear types. Adding some spatial structure to the model should also be considered.*

   This recommendation derives directly from Dr. Dichmont's experience of decapod assessments in Australia. In some of her work, she utilizes a size-based model that runs on a monthly time step to provide management advice. The assessment team recognizes the spirit in which this recommendation is made and believes that it will be of use to explore such a model for the Chesapeake Bay. However, the assessment team also notes that model complexity is in itself not necessary just because it can be done – the Australian example uses in season management tools and thus requires a shorter time step. Were the three Chesapeake jurisdictions to consider in season management for blue crab, then such a model would potentially be required here. However, until that time, the higher resolution assessment model that Dr. Dichmont recommends may not be necessary.

2. *There are clear spatial---temporal changes in the data that have not been fully captured in the assessment. Other crab studies have shown the value of detailed spatial and temporal mining of the data. There would be value in undertaking a detailed and cohesive analysis of spatial and temporal dynamics of different population components on different surveys and their relationship with abiotic factors.*

   The assessment team fully supports this recommendation, but notes that some of these analyses are currently underway through the research awards that supported the assessment. Not specifically mentioned in Dr. Dichmont's recommendation, but one which we believe is also relevant is to conduct similar analyses of the commercial catch and effort data. Resolution of fine scale fishery-independent patterns will only be of use if they are matched with similar scale analyses of the fishery-dependent data. Both categories of analyses would be necessary for assessment model referenced above.

Recommendations

D1. It is unclear whether these different inputs of implicit age are internally consistent. Some evidence of this should be provided.

   This review comment was largely addressed during the Q&A session with the review panel. It related specifically to concerns prior to the assessment meeting on behalf of Dr. Dichmont and the other reviewers over the reliability of the size-age cut off for blue crab. We provided detailed size distributions from the winter dredge survey that

allayed their concerns. This figure is now provided as Fig 3.1 in the assessment document

D2. A possible change in fecundity with respect to size should be monitored in terms of its impact on the assessment.

The assessment team fully support this recommendation. We note that a recent PhD dissertation at ODU provides evidence of such a change and continued monitoring of changes in stock productivity should be monitored.

D3. In the assessment model, fecundity is not weighted by age, whereas it is likely to be a consideration and quite a major factor within the assessment where the category age 1+ implies that all animals greater than age 1 are equally fecund even though 2+ animals are unlikely to breed. Splitting the model into three age classes should be considered.

This is a similar recommendation to A3 above and we refer back to our response to that comment in general.

D4. A research task for future assessments is to study age related fecundity and other related fecundity patterns for Chesapeake Bay.

The assessment team fully supports this recommendation. Given that the population size of blue crab in the Bay has experienced considerable variation over time, the potential exists for density-dependent changes in stock productivity. These should be monitored.

D5. In future, use the tagging data in the assessment to estimate female fishing mortality and natural mortality.

This issue was discussed in depth during the Q&A session. The review panel were made aware that we had indeed tried to include the time series of M derived from tagging directly into the assessment. However, the relatively short duration of this time series and the contradictory nature of the pattern in the data meant that these attempts were not successful – the assessment model effectively averaged through the time series and did not permit interannual variation to be expressed.

However, if the tagging efforts for mature female blue crab are continued, the time series will be very useful in future assessments. The assessment team fully support this recommendation.

D6. In the interim, modify the Brownie model to estimate survivorship and natural mortality through a parameterization that includes M and F, or M and catchability and effort.

The assessment team fully support this recommendation and will work to modify the application of the Brownie model to develop partitioned estimates of vital rates (see comment A5).

D7. Include the Baywide Multispecies Monitoring and Assessment Program in the assessment as a survey index of abundance.

This comment parallels A13 above and we refer back to answer to that recommendation.

D8. Incorporate interaction terms in the second stage of the delta GLM, especially with respect to strata and year, temperature and year for the winter dredge survey. This comment parallels A9 above and we refer back to answer to that recommendation.

D9. Include in the assessment model either using the four other Virginia trawl survey strata as an additional index of abundance or (better option) analyse the Virginia trawl survey and standardize into a single index.

This comment parallels A11 above and we refer back to answer to that recommendation. Dr. Dichmont provides additional specific guidance as to how the VIMS trawl survey data may be standardized.

D10. Conduct more detailed analyses of the dredge survey catchability estimates, especially with regard to availability or else undertake a spatial model of the survey.

This comment parallels A28 above and we refer back to answer to that recommendation. The assessment team fully supports this research recommendation.

D11. Consider applying consistent methods across jurisdictions.

This research recommendation relates to the operation of the winter dredge survey. The survey is conducted by two different agencies – MDNR and VIMS and slightly different approaches are used to how data are adjusted for catchability. This recommendation is a request for more uniformity in how data are collected and analyzed and the assessment team fully supports this recommendation.

D12. The two-stage component of the dredge survey is an extremely valuable component of the analysis and should be continued.

The assessment team supports this recommendation – the continued stratified random survey design with adaptive sampling in areas of high initial abundance should be continued to provide estimates of overwinter mortality. As the time series of relative overwinter mortality expands, it will be of increasing utility in the assessment.

D 13. The dredge survey is essential and should be continued.

The assessment team strongly supports the continuation of the winter dredge survey. It is an essential management tool.

D 14. Run a sensitivity test that uses the raw standardised winter dredge survey indices (not converted to absolute).

See D15 below

D15. Review the derivation of the absolute index of abundance in detail and investigate possible sources of the difference between the absolute and relative model runs (both in terms of the assessment and the survey). This is the highest short---term priority.

During the review, there were several discussions about potential for the expansion of the WDS catches to abundance to cause biases in estimates of sex-specific abundance (see D11). The assessment team fully supports this recommendation and sees it as a potential way to reconcile differences between the sex-specific catch and WDS indices.

D16. Use the daily Potomac data in the assessment, which seem to be very rich and useful.

This comment parallels A16 above and we refer back to answer to that recommendation. The assessment team fully supports this research recommendation.

D17 Independently substantiate these adjusted changes to the catch data through interviews of dealers/fishers or investigate whether there is a discrepancy within the spatial landings data as the fishery effort moves over time.

This comment parallels A15 above and we refer back to answer to that recommendation. The assessment team fully supports this research recommendation.

D18. Undertake model sensitivity tests with unadjusted catch data and an intermediate value of the reporting change factor.

This comment parallels A15 above and we refer back to answer to that recommendation. The assessment team fully supports this research recommendation.

D19. Estimate the reporting change factor (or intervention parameter) internal to the model (best option). If the model is unable to estimate this parameter, then an alternative would be to include the reporting change parameter (with variances) as an input value within the model so sensitivity tests of this parameter can be undertaken and the error in this variable could be included internal to the model.

This comment parallels A15 above and we refer back to answer to that recommendation. The assessment team fully supports this research recommendation.

D20. Investigate the potential of dividing the catch data into gear type or at least undertake a gear analysis.

The assessment team fully supports this research recommendation. Representing the different fishery sectors as separate fleets in the model has a lot of advantages. For example, we could reflect the different spatial and temporal dynamics of each fleet. However, we believe that such an expansion of the number of fleets in the model must be accompanied by a shift to the higher spatial, finer temporal resolution model recommended in Dr. Dichmont's long term research plan. We do not believe it is appropriate to undertake these analyses for the current assessment.

D21. Investigation should be carried out to establish whether it is possible to produce a standardised catch rate dataset that can be used in the assessment, as the index only needs to be representative rather than complete. Also, investigate the use of effort data within the assessment.

This comment parallels A17 above and we refer back to answer to that recommendation. The assessment team fully supports this research recommendation.

D22. Given the value of this fishery in the region, there should be more emphasis placed on collecting accurate and complete catch and effort data that are well aligned with the needs of the assessment.

There have been substantial improvements in the quality of information coming from fishery dependent sources. As noted in response to A17 above, the current assessment effort was amiss in not completing evaluation of the potential of the effort data that is currently available to provide a suitable CPUE index. It is a high priority item for the upcoming period. The assessment team fully supports this research recommendation.

D23. Either record the catch of the soft and peeler crab at the point of landing or record the length of time the crabs are kept in the tank and estimate the survival rate of these crabs.

The research team fully supports this research recommendation. Such data will be vital if the assessment moves toward the high resolution models recommended by this reviewer. Sources of catch and discard mortality in this sector of the fishery are poorly described and improvement in the information available would be beneficial.

D24. The conversion of catch in weight (bushels) to numbers should use the mean weight from the catch for each year both for the past and the future. This means that the mean size in the catch by year is required.

Dr. Dichmont is quite correct in this point, that the approach we took requires information on the mean size of crabs available in the catch. Such data are available in Maryland through their sentinel fishery program. Similar data are not routinely

collected in Virginia. We note that the data available for Maryland are not available for all fishery sectors and may not adequately reflect variation in the sizes of crabs collected in different regions of the Bay. Thus if we are to move to the high resolution assessment model as proposed by this reviewer, it is important that an adequate baywide sampling program be established to monitor the biological characteristics of the harvest. The assessment team fully supports this recommendation.

D 25. Undertake a baywide recreational survey as the resource is increasing and recreational effort may be increasing. This is a priority.

This comment parallels A18 above and we refer back to that recommendation. The assessment team fully supports this research recommendation.

D 26. Undertake sensitivity tests for larger recreational catches in the assessment.

The model utilized an 8% correction in landings to reflect the magnitude of the recreational catch. As noted in Section 5.4.2, the magnitude of the recreational catch is poorly described. As such, it could be argued that we should have explored the sensitivity of the model to this source of uncertainty. However, because individual fishery sectors are not recognized in the model – that is harvest occurs as a result of the actions of single composite fishery, the assessment team believes that sensitivity analyses to specifically examine the consequence of uncertainty in the recreational fishery as a single sector are not warranted. Because the model includes recreational catch as a constant proportion of the commercial catch, the effect of changing the assumed 8% is likely a proportional change in abundance (as is often seen in other assessment models).

We believe that a broader analyses of the sensitivity of the model to misreporting generally would be more appropriate and we did not undertake such an analysis. The assessment team recommend that analyses of the potential impact of misreporting on reference points and stock status be a high priority.

D 27. Undertake full sensitivity tests of the stock---recruitment relationship (following this into the reference point calculations) and also develop a function that captures the principle of what is presently formulated but does not optimise at zero males i.e. incorporate a compensatory function in the stock---recruitment relationship.

There remains considerable disagreement between the assessment team and the review team regarding the stock-recruitment model used in the assessment. The review team point out, quite correctly, that when taken to the extreme the stock recruitment model used in the assessment predicts continued recruitment when there are no males in the populations. As Dr. Dichmont notes, the model predicts the best thing to do is to catch all the males! This is clearly not biologically reasonable. We revised the assessment report to strongly caution that managers should implement policies to stay

within the empirically observed sex ratios – and that pursuing policies that cause the population to move to more extreme sex ratios is fraught with potential danger of overfishing males and thus reducing the stock's productivity.

In its place, the review panel recommended we implement a standard non sex-specific stock recruitment model. We have implemented this in additional sensitivity runs included in the assessment document and Assessment Working Paper III. We note that the uMSY reference point from this combined sex stock recruitment model is identical to that developed in the sex-specific base model. We further note that the non sex-specific stock recruitment model recommended by the review panel is not free of absurd implications in the face of extreme sex ratios. Specifically, because the sex composition of the spawning stock is not specified in the non sex-specific stock recruitment model, the implicit implication that average recruitment will be observed from stocks comprised of a normal sex ratio, an all female population and an **ALL** male population.

The assessment team continues to believe that the sex-specific stock recruitment model used in the assessment is more appropriate for blue crab than the non sex-specific model. To address the review panel's concerns we are currently working on a modified version of the sex-specific model that imposes a recruitment penalty on increasing female dominated populations to reflect sperm limitation. However this refinement of the model is not yet fully developed.

D28. Develop a conceptual model which would provide a much better description of the biological timeline and how this connects to the assessment.

This comment parallels A22 above and we refer back to answer to that recommendation. We have included an additional figure to reflect these changes.

D29. The decrementation in, for example, equations 7 and 11 etc. should index M with time of year and F by effort.

This comment parallels A23 above and we refer back to that recommendation.

D30. Split the partial recruitment parameter (that ties selectivity and growth) to include gear specific selectivity. The model should also overtly include selectivity of the gear (including how it changed over time).

The assessment team notes that the current assessment model does not represent separate fishery fleets in the model, but rather models removals as occurring from a single composite fishery. As a result this recommendation relates to Dr. Dichmont's long term recommendation that a size-based, high spatial resolution, fine temporal resolution assessment model be developed. Were such a model to be developed, then this research recommendation becomes not only relevant but imperative.

D31. For consistency, set the initial conditions correctly allied with the assessment implementation i.e. from 2 to 4 parameters.

This comment refers to assumptions made in the model about abundance in the four sex-age categories in the first year. The model currently estimates one parameter for the initial number of total recruits and another for the initial number of males and females. The total recruits are partitioned into males and females based on the observed sex ratio from the WDS (the same assumption that is used to partition recruitment for all later years). In contrast, no difference in sex ratio of fishing mortality rates is applied for abundance in the age-1+ category. Thus, the model assumes that abundance of age-1+ males and females is equal in the first year. The assessment team agrees that it would make more sense to estimate separate parameters for initial age-1+ abundance of males and females and that this should be pursued in the next assessment.

D32. Include the calculation of effective sample size for the winter dredge survey in the likelihood of the assessment.

We believe that this recommendation refers to the methods used to obtain the weights applied to the different fishery independent data sources in the model. The model uses a weighting approach where the fit of each index of abundance is weighted inversely with its standard deviation (SD; or log-scale SD). The SDs of the likelihood components of the three surveys used in the assessment were estimated differently. The likelihood components for the MD and VIMS trawls were estimated by iteratively changing the assumed SDs until the input SDs were approximately equal to the output SDs for each survey. The approach of estimating the SDs of the indices of abundance is often used when estimates of precision from the indices are not thought to fully represent the uncertainty in the index. However, the WDS was treated differently. Because of the large number of samples, the full sampling of the stock's range, and the high degree of interannual consistency in the distributions of catch each year, we used the design-based estimate of the variance in the likelihood function for the winter dredge survey. This essentially assumes that the survey design fully describes the uncertainty in estimated abundance. This does not require any correction for effective sampling size. The assessment team recommends that in the future simulations be conducted to assessment potential differences between results from using both approaches.

D33. Include the bias correction in the stock---recruitment relationship.

This comment parallels A23 above and we refer back to that recommendation.

D34. There would be some benefit in running sensitivity tests where the model is formulated as either process or observation error driven.

The model currently includes both observation error in the indices of abundance and process error in the catch data. The assessment team agrees that this is an important research recommendation that will help us understand more fully the tension that model experiences in trying to fit the survey and catch time series.

D35. Undertake a sensitivity test of the influence of the different indices as well as implement the dredge survey as a relative index.

The assessment team has undertaken specific additional sensitivity analyses that are included in the assessment document (Table 6.2) and in Assessment Working Paper III. These sensitivity analyses explored the impact of treating the winter dredge survey estimates of abundance as a relative index, rather than as an absolute index as currently used. The overall effect of assuming that the winter dredge is a relative index is to suggest higher reference points than are developed in the base model. We have included these results in the report, but have retained the base model because we believe it is a more accurate representation of the WDS and because it is more precautionary than the models that assume the winter dredge survey is a relative index.

D36. Run sensitivity tests of different input data and parameters and its effects on reference points. Consider systematic approaches such as FAST or a designed experiment.

The assessment team agrees that additional exploration of the sensitivity of recommended reference points to uncertainty in model inputs is required. It was an original goal of the assessment team to bring forward not only point estimates of the reference points, but also an indication of the uncertainty associated with each. We are committed to continuing work in this area and fully support the research recommendation.

D37. Provide more detailed fit statistics like q---q plots and residual plots for all indices.

We have added an additional index to the assessment document that now provides all outputs from the base run of the model. We provide individual annual estimates for each of the time series that were fit in the model and for key population parameters.

D38. Provide more choices of reference points rather than providing single values for each.

This comment parallels A24 above and we refer back to answer to that recommendation. In order for managers to make use of information on multiple reference points, additional research would need to be conducted to characterize the tradeoffs inherent in adopting a set of reference points and a harvest control rule. The assessment team fully supports this recommendation.

D39. Provide the option for more precautionary limit reference points.

The assessment team and the review panel spent considerable time in the Q&A session over the management philosophies in different regions. There are clearly no absolutes when it comes to reference points – with the exception perhaps of MSY – are to an extent arbitrary. For example SPR reference points are "more guidelines than a code." The assessment team followed the tradition that has developed in the Chesapeake region and in the mid-Atlantic generally of using MSY and 0.5*MSY as reference points. The review panel is quite correct to point out that there is no a priori reason why these management points should be selected. Indeed the assessment team strongly believes that target reference points should be negotiated among managers and stakeholders to reflect societal decisions regarding the best use of the resource. Specifically target reference points are not a scientific choice, but a socio-political one. Science can provide values following guidelines, but should not be the sole arbiter of the decision making.

The assessment team fully supports this recommendation.

D 40. The assessment document should be updated to include various corrections and clarifications as described in the review report.

The assessment team has made all of the changes recommended by Dr. Dichmont that relate to information not provided, or incorrect references within the document.

Dr. Billy Ernst.

Dr. Ernest provided 24 recommendations that were of both a short term and a longer term nature. Solely because Dr. Ernst's review is the third one we discuss, many of his comments have already been raised by the two previous reviewers. We detail our response to each of his comments below.

1. There is an appropriate description and mathematical representation of the stock assessment model in the main document; nevertheless it is necessary that the authors include some modifications in the final document:
a. A conceptual model should be included, which would provide a better description of the components, biological timeline and how this connects to the assessment, especially with regard to survey data. Main model assumptions should also be addressed.

We have addressed this concern with additional text in the assessment document.

b. The stock recruitment function should include the bias correction factor to avoid misinterpretation of recruitment parameters estimates.

This comment parallels A23 and D33 above. We refer back to these responses in addressing this comment.

c. Initial conditions of the population dynamic model should be clearly specified in the document.

We have addressed this concern with additional text in the assessment document (see also comment D31).

d. Appropriate documentation should be provided of the non-precautionary nature of the stock-recruitment function, particularly at low population stock size.

We have included additional text regarding the reviewers concerns over the stock recruitment model. We continue to defend robustly our choice of stock-recruitment function, but have including cautionary notes to managers in the assessment document regarding the risk of exceeding the observed ratios of sex-specific exploitation rates.

e. A better description is needed of the sensitivity tests presented in the document, including estimated parameter values and standard errors, likelihood values associated to each data component, and uncertainty around the main model outputs.
f. Sensitivity tests undertaken during the review should be described.

All sensitivity runs are now fully documented in Table 6.2 and in Assessment Working Paper III. Additionally we provide all output of the base simulation in Appendix V to the assessment document

g. Data points on the reference point figures (6.1 and 6.14) should be annotated with the years, as this would link better with the text.

We have changed the two reference point figures as requested.

h. A composite reference point figure should be provided, including: a) the 2005 method for standardizing the survey data and the 2005 reference points, b) the 2010 standardized data with the 2005 reference points and c) the 2010 standardized data with the 2010 reference points. This figure should be accompanied with an interpretation of the changes when moving from the 2005 to the 2010 assessment.

We have changed the assessment document as requested.

2. In many ways this fishery can be considered data rich and from all available data the winter dredge survey is the most important piece of information. It is used in the assessment and the evaluation of stock status relative to the reference points. It is important that this survey is continued.

The assessment team fully supports this recommendation. The WDS is essential to the ongoing management of the blue crab fishery in the Chesapeake Bay. This survey is the single highest priority for funding.

The following four points need no response

3. The structure of model (Age-0/Age-1+ and sex) seems appropriate considering the nature of the historical data. It also allows one to keep track of meaningful population statistics (spawning abundance by sex), which is of special relevance under recent changes to the management regime in the fishery.

4. The calculation of reference points is incorporated within the assessment model. This is a major advantage over previous models, because it assures consistency between statistics.

5. One inconvenience of collapsing the entire adult age structure into one age group is losing track of female fecundity at age. The implicit assumption is that female fecundity remains constant, which is probably rather unlikely to happen. Additionally it is unclear how this would influence the spawning stock size at MSY calculations.

6. In the assessment report there was an important effort to assess model structure uncertainty (results of two additional models were presented). Despite the intrinsic value of this exercise, a direct comparison of reference points between the models is not possible, because female spawning stock is not available for the other models. Nevertheless of the three assessment models provided, the best model is the SSCMSA.

7. On the other hand the parameter uncertainty of the proposed model was not fully characterized through the documented sensitivity tests, nor was the robustness of reference points estimates adequately presented. For example, test runs during the review highlight that the assessment model output is extremely sensitive to treating the winter dredge survey as a relative index of abundance. The fit was much better than the base case presented in the report. This is of great concern, as traditionally the default assumption is to treat surveys as a relative index of abundance. It is recommended that:

a) A sensitivity test is run using the raw winter dredge survey indices.
b) The entire calculation of survey catchability should be checked in detail and possible sources of difference should be investigated.

The assessment team has completed the additional sensitivity run requested and it is now included in the assessment document.

The assessment team remains concerned over the estimation of survey catchability for the winter dredge survey. There remain differences in survey protocols between the two jurisdictions. This is particularly unfortunate given that the ecology of blue crab is believed to differ along the north-south axis of the Chesapeake Bay – thus determining to what extent survey inferred difference relate to differences in the way the survey is conducted and the data are handled between the two states or to differences in the underlying ecology is difficult. The assessment team strongly recommends that an interjurisdictional team (possibly CBSAC) be assembled to review the operation of the WDS, the theoretical and practical foundation for the calculation of survey catchabilities and intergration of the two data streams. The assessment team believes this is of the highest possible priority.

8. The reference point calculations seem appropriate based on the structure of the new model. The authors followed the Federal reference points system to define the limit reference points and the New England and Mid-Atlantic Fishery Management Councils' system to define target reference points. This seems appropriate, but is in the end a policy decision. With respect to stock based reference points there is a value judgment about how precautionary the overfished limit and target reference point needs to be. Different values arise using different male to female F ratios, and coincidentally the one chosen in the report is not the most precautionary one.

Future work

9. This assessment lacks a catch per unit effort index of abundance, an index that captures the condition of the stock from the fisheries perspective. Future assessments should investigate the feasibility of generating a standardized catch rate index. This would ultimately be of importance in future economic analyses.

This recommendation parallels similar recommendations from the other reviewers (A17 and D21, D22). The assessment team fully supports this recommendation. We believe that several potential approaches might be fruitful in this regard:

a) analysis of catch and effort from the Potomac River Fisheries Commission – the PRFC has maintained accurate, spatially resolved catch and effort data in the Potomac River since 1963. CPUEs are routinely calculated from these data and provided to the Commission and its workgroups. However, these data have not been utilized within the assessment as they should have been.

b) Both states are now collecting mandatory daily reports from individual waterman. Although the time series is, at present, of insufficient duration to be of direct use in the assessment currently, that is not likely to be the case when the next assessment is conducted. Efforts should begin now to assemble and review these data

c) While catch and effort data from the two states may be too short to be of direct use in an assessment currently, they do provide extensive spatial information on the pattern of fishing on a daily and seasonal basis. It is likely that, were these data combined with data from fishery-independent surveys considerable insight could be gained in the response of the fishery to changes in the underlying distribution of blue crab.

The remaining 14 points are fully covered by responses to the comments of the two other reviewers – in our responses we refer back to our specific comments

10. The substantial difference between the model and estimated catchability coefficient of the winter dredge survey requires immediate attention. This is a high priority task.

This comment parallels A9 above and we refer back to answer to that recommendation.

11. The spatial component seems to play an important role in the population dynamics of different ontogenetic stages of Chesapeake Bay blue crab. A detailed data driven analysis (combining all of the pieces of information) of spatial and temporal dynamics of different stock components should be undertaken. Environmental covariates should also be factored in.

This comment parallels A7 and Dr. Dichmont's second long term priority above and we refer back to answer to that recommendation.

12. More work on the natural mortality parameter estimation should be undertaken using the available mark recapture data. Alternative Brownie model parameterization should be considered in the analysis, to explicitly model natural and instantaneous fishing mortality rates. Ultimately, the tagging data could be directly incorporated into a more integrated stock assessment estimation model.

This comment parallels A5 and D6 above and we refer back to answer to that recommendation.

13. Only a nominal index of abundance was constructed from the Virginia trawl survey data during this assessment. Index standardization should be undertaken. A second index of abundance can be potentially developed based on the winter portion of the Virginia trawl survey, which could provide information on the female spawning stock.

This comment parallels A11 and D9 above and we refer back to answer to that recommendation.

14. Interaction terms were omitted in the standardization of survey indices. It is recommended to develop additional models with interaction terms in the second stage of the delta GLM analysis.

This comment parallels A9 and D8 above and we refer back to answer to that recommendation.

15. In recent years some areas in the Chesapeake Bay seem to be experiencing higher recreational fishing mortality than reported in this analysis. It is recommended that a new baywide recreational survey be undertaken.

This comment parallels A18 and D25 above and we refer back to answer to that recommendation.

16. Despite the data-rich nature of this assessment, there is a major need for accurate and complete catch and effort information. Some of these factors are presented in the following list:

(a) the conversion of the catch in weight to numbers should be based on measured average weight in the catch, rather than using a constant factor. This means that mean size in the catch by year is required;

(b) a better and more direct estimate of soft and peeler crab mortality should be developed for the assessment;
(c) the adjustment of past changes in catch reporting requires some independent verification. The catch data is influential to the assessment results; (d) Anecdotal information suggests that incidental winter dredge and summer pot fishery mortality might be important. It is important to quantify these additional sources of mortality.

This comment parallels D22-24 above and we refer back to answer to that recommendation.

17. Additional model runs should be undertaken to assess the robustness of the results to:
(a) using the modified/original time series of catch;
(b) estimating the catchability coefficient of winter dredge survey for adult crab;
(c) using different pieces of information;
(d) not estimating observation and process error simultaneously;
(e) to free up the assessment model to allow for a better fit of extreme values. A sensitivity test might be organized in a systematic and comprehensive way, using appropriate techniques (i.e. FAST, designed experiments).

This comment parallels A26-28 and D34-36 above and we refer back to answer to that recommendation.

18. Include additional sources of information into the assessment such as Baywide Multispecies Monitoring and Assessment Program and Potomac daily catch data.

This comment parallels A13 and 9 above and we refer back to answer to that recommendation.

19. Adjust the reporting changes in the catch internal to the assessment model and input the intervention parameter so that this acknowledges some of the uncertainty.

This comment parallels A9 above and we refer back to answer to that recommendation.

20. Incorporate formally the different fisheries into the assessment. In addition reparameterize the partial recruitment parameter into formal selectivity and growth components.

This comment parallels A2 above and we refer back to answer to that recommendation.

21. The initial conditions of the model should have 4 as opposed to 2 parameters. If initial conditions stay in a 2-parameter configuration, the implicit assumption for the other 2 parameters should be consistent with the rest of the model configuration.

This comment parallels A9 above and we refer back to answer to that recommendation.

22. In equation 7 and 11, kappa should be multiplied by only F and M should be scaled by time.

This comment parallels A9 above and we refer back to answer to that recommendation.

23. Incorporate some kind of compensatory component into the stock recruitment relationship to avoid non-precautionary behavior of the model (i.e., depleting male population, while maintaining high recruitment rate).

This comment parallels A9 above and we refer back to answer to that recommendation.

24. A more complex version of the model might include:

(a) size structure;
(b) monthly time step to allow for the evaluation of in-season regulations; and
(c) spatial structure.
The model development could be frameworked into a MSE type of configuration.

This comment parallels Dr. Dichmont's first long term priority recommendation above and we refer back to answer to that recommendation.