

# Summary Review of the Chesapeake Bay Blue Crab (*Callinectes sapidus*) 2005 Stock Assessment.

**Malcolm Haddon (Compiler)**

Marine Research Laboratory, Tasmanian Aquaculture and Fisheries Institute  
University of Tasmania  
Private Bag 49, Hobart, TAS 7001  
Australia  
[Malcolm.Haddon@utas.edu.au](mailto:Malcolm.Haddon@utas.edu.au)

**Contributions by Michael Bell, Nick Caputi, Malcolm Haddon, and Paul Medley.**

Executive Summary.....	2
Recommendations .....	3
Background.....	9
Statement and History of the Problem.....	9
Review Activities .....	9
Summary of Findings .....	12
Structure of Document .....	12
Terms of Reference .....	12
Evaluate the adequacy and appropriateness of all data used in the assessment, including the following: .....	12
Life History and Vital Rates of Blue Crab in Chesapeake Bay.....	12
Patterns in Fishery-Independent Surveys. ....	16
Patterns in Catch and Effort by Sector and Region. ....	20
Evaluate the adequacy, appropriateness, and application of the assessment models used for the Chesapeake Bay blue crab fisheries and characterize the uncertainty in the assessment. ....	22
Evaluate the scientific basis for the control rule for the Chesapeake Bay blue crab fishery. ....	26
Develop recommendations for future research for improving data collection and the Chesapeake Bay blue crab assessment. ....	28
Conclusions .....	28
Bibliography .....	29
Appendix 1: STATEMENT OF WORK .....	32
Appendix 2: Description of Chair’s Summary Report .....	35
Appendix 3: Assessment Model Analysis .....	36
Appendix 4. GLMS for Depletion Experiments.....	40

## Executive Summary

The most important commercial fishery within the Chesapeake Bay targets the blue crab (*Callinectes sapidus*). Commercial landings of blue crab have exceeded 100 million pounds historically (1993) with more recent average landings reaching approximately 72 million pounds. The economic value of the blue crab fishery to the Chesapeake region exceeds \$200 million annually.

This fishery has now been assessed two times; the first time was by Rugolo *et al.* (1997) while the most recent (Miller *et al.*, 2005) has only just been produced. Because of the importance of the blue crab fishery to the region the Chesapeake Bay office of NOAA requested a formal review of the latest assessment document and the work behind it. To that end the Centre for Independent Experts, organized out of the University of Miami, arranged for four reviewers to attend a review workshop held in Annapolis over the period 9<sup>th</sup>-11<sup>th</sup> August, 2005. The NOAA office provided a large amount of written material and data on a CD before the review meeting. The review aimed to examine the validity and appropriateness of the data used in the assessment, the model and its structure, the management decision rules (control rules) that flowed from the assessment, and also requested suggestions for further research that would benefit future assessments. The review was conducted in an open and friendly atmosphere, which was maintained throughout even the most intense discussion.

The 2005 assessment was found to involve significant improvements over the previous 1997 assessment document in all areas that were reviewed. Biological properties critical to an appropriate and valid assessment include the estimates of natural mortality, crab growth, and recruitment. Significant changes and improvements in all three areas have been made and in many cases published. Thus, natural mortality was addressed by Hewitt & Hoenig (2005), estimates and studies on individual growth have been produced by Bunnell & Miller (in press), Miller and Smith (2003), and Ju *et al.* (2001). Reproduction and recruitment issues have been considered by Hines *et al.* (2003) as well as by the continued surveys conducted by the Virginia Institute of Marine Science (Lipcius & Stockhausen, 2002; Lipcius *et al.*, 2003). In addition to improvements stemming from biological studies the reporting of current catch has been improved with new regulations, the instigation of an observer program, and recreational surveys. These have been complemented by the standardization of the previous catch history through a detailed statistical study that attempted to account for changes in reporting methods that have occurred through time (Fogarty & Miller, 2004). The single most valuable improvement has been the maturation and development of the winter dredge survey, which, because of its geographical coverage and precision will provide the critical information necessary for management into the future. The methods used are now standard with the use of depletion experiments across strata and substrates to estimate catchability and improve this survey as an estimate of absolute abundance (Volstad *et al.*, 2000; Sharov *et al.*, 2003). The stock assessment model has been extended to use a number of separate survey abundance indices as data, as well as estimates of exploitation rate deriving from the winter dredge survey. This latter innovation extended to the use of the exploitation rate (fraction of available biomass taken in the fishery) in some revised control rules to aid in the management of the fishery. The management of the fishery has also been enhanced through the implementation of closed areas in Virginia aimed at securing a larger proportion of the spawning females (Lipcius *et al.* 2003).

The model was an improvement over the previous model through its use of more data but was found to be relatively unstable and it was suggested that this may stem from the still limited amounts of data available compared with the relatively large number of model parameters being estimated. It was suggested that the same model be examined but using only observation errors and not the process errors. This would lead to a more stable model and possibly to somewhat different management implications.

The control rules developed for this fishery appear well structured and well defined (though with a degree of arbitrariness as to the exact levels selected as limits and targets). Unfortunately they are not supported by a well developed set of management outcomes should the control rules be triggered. It is recommended that clear and explicit objectives be determined for this fishery (this is a job for the managers not the scientists) so that suitable and acceptable management responses can be developed to respond to situations where the stock is found to be in an undesirable state. Some way of limiting the apparently run-away effort currently being expressed is urgently required.

There is a possibility that this fishery is more strongly influenced by recruitment events than by mortality events (fishing). If this is the case it will make management more difficult.

## **Recommendations**

The recommendations below are structured approximately in line with the terms of reference of the review. They derive from the four separate panellist's reviews and are not given in any particular order under each heading. Where similar recommendations were produced by different panellists the emphasis given to each recommendation often differed among reviewers. The progress made between the two assessments was large and comprehensive. However, with the uncertainties in the available fisheries data, the large latent effort, and the unprecedented current efforts levels, there remain many things that could be done to improve the assessment. All reviewers agreed with the recommendations already in the assessment document. Those, combined with those listed below should provide the basis for constructing a strategic research plan for the Chesapeake Bay blue crab fishery assessment.

### **Life History and Vital Rates of Blue Crab in Chesapeake Bay**

- 1) The significant contribution of cannibalism to natural mortality suggests that  $M$  could be inversely related to age/size because of reduced levels of cannibalism and other predation with increasing size and a reduction in frequency of moulting. It is recommended that the possibility of a size-related difference in natural mortality be examined. At the very least, sensitivity trials with the assessment model using plausible schedules of such non-constant natural mortality should be continued and expanded. This could have marked effects upon the outputs from the modelling.
- 2) It is recommended that the possibility of some density dependent relationship between natural mortality and population density be explored. Such a relationship seems intuitively likely if cannibalism is a major source of mortality in the younger crabs.
- 3) It is recommended that the environmental drivers behind the growth transition from undersized crabs to legal sized crabs be examined in more detail. Within year dynamics are largely ignored in the assessment but this growth would have

important implications for availability through the year. An assessment of the sensitivity of the model to some annual differences in the size threshold is required.

- 4) It is recommended that the possibility of density dependent changes in growth with population density be examined. Such changes would significantly alter the stock productivity.
- 5) One of the fastest growth models used to obtain the average growth model used in the stock assessment model was derived from a pond mesocosm (Ju et al. 2001), which may not be representative of growth in the wild. It is recommended that the sensitivity of the assessment results to the removal of this extreme value be documented.

#### **Patterns in Fishery-Independent Surveys.**

- 6) Fishery-independent survey data provide an excellent resource for assessment of blue crab stocks. It is recommended that the spatial, temporal and size criteria used to define abundance indices be re-examined. In particular, there may be scope to improve the separation of pre-recruits from the fully recruited stock.
- 7) It is recommended that the spatial information from the winter dredge survey be used in conjunction with time series of environmental conditions to examine any relationship between recruitment (as 0+ abundance) and environmental drivers such as temperature and salinity. The spatial detail in the other surveys should also be investigated as, combined, they may provide an avenue to throw light on whether environmental variation or urban development is related in any way to trends in recruitment abundance for each sex separately.
- 8) It is strongly recommended that the winter dredge survey continue to be conducted in its present form. It will provide the necessary estimates of abundance required into the future. It is recommended, however, that confidence intervals be estimated for the exploitation fractions, incorporating uncertainty in the estimation of abundance (which includes uncertainty around the mean CPUE and uncertainty in the catchability values applied to convert CPUE to density) and uncertainty in the estimates of numbers in the catch.
- 9) It is recommended that the depletion experiments used in the winter dredge survey to estimate catchability in different strata be improved by using differential GPS and associated plotters to mark the depletion tracks. This along with standard data collection of sea conditions and gear should enable the use of Generalized Linear Modelling to improve the analysis of the experiments.
- 10) It is recommended that Generalized Linear Models be used to improve the analysis of the Maryland and Virginia trawl surveys. By standardizing these indices of relative abundance, confounding factors will be removed from consideration and a cleaner signal should be presented. In particular it could provide corrections for changes in gear, weather, habitat and survey design that might have occurred during the years. This should provide better indices of relative abundance for both the 0+ and 1+ cohorts.
- 11) The Virginia trawl survey provides a workable index of mature females on the spawning grounds. This time series of spawning female abundance should continue to be monitored closely and could be considered as a source of information for a possible Biological Reference Point for this fishery.

- 12) It would be useful to explore an additional abundance index made up of a recruitment index from as early in the life history as possible. This will increase the predictive power of any stock assessment model and improve advice and management planning. The best candidate for this would be the current megalopal survey, but the Virginia trawl survey index of spawning females would be an alternative. This series may assist in the examination of the impact of environmental conditions on recruitment.
- 13) The Virginia spawning sanctuary and protected deepwater dispersal corridor has been used successfully to secure the level of spawning stock. The closures appear to be a workable management option for securing and possibly enhancing the level of the spawning stock and further research is required into improving the spatial and temporal spawning closures.
- 14) It is recommended that a formal statistical study be conducted to characterize the relative contribution to the assessment of each of the time series of survey estimates. Of particular importance is the peak of abundance observed in the Virginia trawl survey in 1970 and 1971, which influences the apparent productivity of the complete stock.
- 15) It is recommended that a formal study be made of correlations between the different survey time series to highlight any consistencies or inconsistencies between the surveys. Their use and value in the assessment modelling could be enhanced through knowing about their inter-relationships. In addition, the abundance of the year classes for the WDS could be obtained for Virginia and Maryland separately to assess the relationship between the WDS and the trawl surveys for each state. Finally, the relative weights applied when fitting each time series may be able to be modified. Because of its greater accuracy and generality more weight should be given to the winter dredge survey results.
- 16) Rather than consider the different time series separately, it would be possible to use a GLM to combine the surveys (except the WDS) to produce a single index of relative abundance through time. This single index should be scaled to match the WDS over their period of overlap, thereby providing a more defensible extension backwards in time.
- 17) The inverse relationship between exploitation fraction and abundance, which is suggestive of depensatory behaviour, needs further exploration. The factors operating in the fishery that could induce such depensatory behaviour should be examined. They could include a concentration of the stock during periods of lower abundance or a change in the targeting ability of fishers at different stock sizes. The impact of variability on the abundance estimate may bias this relationship and this should be examined. A density-dependent effect on natural mortality could also have an impact on this relationship and should be investigated using sensitivity analysis. Finally, the correlation could simply be a statistical artefact, arising from the use of stock abundance to calculate exploitation fraction. It is recommended that a randomization test or other appropriate method be used to characterize the contribution to the correlation made by this arithmetic relationship.

#### **Patterns in Catch and Effort by Sector and Region**

- 18) It is recommended that the commercial catch and effort data collections be standardized, if possible, between Virginia and Maryland, making it easier to share the data. A central repository of the combined data should be easily available to

enhance the opportunities for assessment. Efforts should be made to obtain more accurate information concerning the currently under-reported peeler fishery.

- 19) It is recommended that a review of regulations across the two States be undertaken with an aim of increasing commonality between the two States, which in turn may improve the ability to collect consistent data.
- 20) The time-series methodology used to correct commercial landings with respect to reporting changes are statistically rigorous, and provide the most scientifically defensible estimates of trends in fishery removals. It is recommended that the statistical uncertainty involved in this correction be incorporated into measurements of uncertainty for derived values such as catch numbers and exploitation fractions.
- 21) Catch numbers are estimated from catch weights assuming that mean sizes in fishery-independent surveys are representative of commercial removals. This assumption should be validated by sampling of commercial landings.
- 22) It is recommended that experiments and observations be designed to investigate the potential mortality arising from discarding undersized crabs from commercial pots.
- 23) It is recommended that experiments and observations be designed to investigate the potential mortality arising through the agency of ghost fishing by self baiting lost crab pots. If this is found to be significant then options for ameliorating the problem should be developed.
- 24) It is recommended that a study be made of the incidental mortality induced by the different fishing methods used for blue crabs. Once a reliable estimate of the amount of effort by gear type is available this will provide another estimate of fishery related mortality.
- 25) It is recommended that recreational fishing continue to be monitored and the use of licences be encouraged. Recreational fishers using trotlines require a licence and the number of licences sold each year should be monitored as an indicator of the trend in effort. A random sample from this group could also be selected to undertake a phone/diary survey of their fishing activities during the year.
- 26) A cross check on the accuracy of the catch estimates is being examined in Maryland using dealers' catch reporting and voluntary daily reporting of catch rates by selected cooperating fishers and observers on board commercial vessels. This work should be continued and extended into Virginia.
- 27) With the implementation of the compulsory commercial fishery returns, the effort data needs to be examined. This is important in the assessment of nominal effort trends, catch rates and evaluation of issues such as pot saturation and latent effort. The fishery-independent spatial survey of effort of the trap fishery in Maryland should be continued to help interpret and validate the fishers' returns.
- 28) The latent effort in the blue crab fishery needs to be assessed in terms of numbers not fishing, number of part-time fishers, and periods of low fishing activities amongst full-time fishers, as this will affect any attempt to try and reduce fishing effort.

**Evaluate the adequacy, appropriateness, and application of the assessment models used for the Chesapeake Bay blue crab fisheries and characterize the uncertainty in the assessment.**

- 29) It is recommended that alternative configurations of the assessment model be considered, especially simplifying the model by excluding the use of process errors. This will significantly reduce the number of estimable parameters and should stabilize the model and permit the inclusion of the exploitation fraction in the objective function used to fit the model.
- 30) The uncertainties in the data and analyses are not included alongside the assessment outcomes. It is recommended that bootstrapping or other methods be used to provide confidence intervals around the assessment estimates. With a simplified model this would be simpler to implement. The inclusion of uncertainty would allow a risk assessment when comparing assessment outcomes with biological reference points.
- 31) Given the high estimated mortality rates, significant benefits may accrue through switching to shorter time steps than a year; which would allow a reasonable chance of estimating fishing mortality. The first step would be to adapt the current annual stock assessment to a within-year model. This would mean describing within season processes, such as the time of arrival of recruits, and generating data based on the shorter time step (e.g., catches in each month). The shorter time step will help model the population dynamics of an animal with such a short life history and measure the impact of fishing.
- 32) The stock assessment does not take into account the possible differences in biology and exposure to mortality between male and female blue crabs. It is recommended that at least the sensitivity of the assessment to sex-related differences in exploitation rates be examined, and the possibility of modifying the assessment model to include the sexes explicitly be explored. If there is a higher mortality on females a separate treatment may identify a greater impact on spawning potential than a combined assessment.
- 33) The documentation of the model in the assessment document should be expanded to include the details of all equations along with a comparison/justification of the errors structures used, demonstrating the advantages or otherwise of including process errors in the modelling.
- 34) Biological Reference Points (BRPs) are defined for exploitation fraction, aimed at preserving 10% (limit) or 20% (target) of the maximum spawning potential. An individual-based per-recruit model was used to calculate these BRPs. Both the assessment estimates of exploitation fractions and the BRP estimates against which they are compared are very sensitive to assumptions about  $M$ . It is recommended that all sources of uncertainty around the comparison with BRPs be considered in the context of risk assessment.
- 35) A preliminary economic assessment of the fishery should be undertaken in conjunction with the stock assessment modelling to assess ways to improve the economic performance of the fishery. This should provide an assessment as to whether the number of pots being used can be reduced without much impact on the catch being taken. This may result in considerable cost savings with little or no loss of catch.

- 36) It is recommended that the use of short- and medium-term stock projections be investigated with a view to determining likely stock trajectories in relation to minimum abundance thresholds.

**Evaluate the scientific basis for the control rule for the Chesapeake Bay blue crab fishery.**

- 37) A 3-year moving average of the exploitation rate and the abundance could be used as part of the control rules to assess the trends and avoid the annual fluctuations. The annual indicators of catch and abundance can be affected by environmental factors that can both positively or negatively affect catchability. Therefore an average over 3 years will avoid the short-term impact of catchability changes and assist in focusing the control rules on the significant trend in the fisheries.
- 38) To generate fully defensible decision rules regarding the spawning potential, a stock recruitment relationship is required. This could be based on historical recruitment, either observed or modelled. It should be possible to define the point at which recruitment overfishing occurs based on such a stock recruitment relationship.
- 39) An exploration should be made of decision rules based on total catch. As the winter dredge survey (WDS) is strongly correlated to catch, management decisions can be based either on each WDS result or on some simple derivative, such as a moving average value. This approach may be robust to whether the fishery is influenced more by recruitment than by fishing mortality.
- 40) It is recommended that clear and explicit objectives be determined for this fishery (this is a job for the managers not the scientists) so that suitable and acceptable management responses can be developed to respond to situations where the stock is found to be in an undesirable state. The current fishery 'target' is more akin to a precautionary higher operational limit, reflecting uncertainty about real values and acceptable risk. Comparison of stock and exploitation indices with BRPs should be placed firmly in the context of uncertainty and risk assessment. A definition of acceptable risk is needed from managers for this to be applied in practice.
- 41) There is a strong possibility that recruitment rather than mortality is the driving force behind blue crab population dynamics in Chesapeake Bay. If true, this would have strong implications for fishery management in that control of fishing mortality would not be an effective management tool. It is recommended that:
- a. The fishery-independent survey indices be examined (e.g., by GLM) to identify the extent to which recruitment or mortality are the driving forces behind inter-annual fluctuations;
  - b. The fishery-independent survey indices and assessment outputs be examined to determine whether a stock-recruitment relationship can be described; and
  - c. Monte Carlo simulations be used to examine the performance of the BRPs in preventing adverse stock trends.

## Background

### Statement and History of the Problem

In the Chesapeake Bay the most important commercial fishery targets the blue crab (*Callinectes sapidus*). Commercial landings of blue crab have exceeded 100 million pounds historically (1993) with more recent average landings reaching approximately 72 million pounds. The economic value of the blue crab fishery to the Chesapeake region exceeds \$200 million annually.

Sound management of this resource requires accurate information on the status and trends of the blue crab population and on the dynamics of the fisheries that exploit the stock. There have been two recent stock assessments completed for the blue crab (Rugolo *et al.*, 1997; and the document under review, Miller *et al.*, 2005). In addition, the NOAA Chesapeake Bay Office (NCBO) has produced annual 'Advisory Reports' for blue crab to assist resource managers in the decision making process.

The 2005 assessment here under review was initiated in October 2003. Due to the political nature of any decision regarding fisheries in Chesapeake Bay, especially that for blue crab, an independent and expert review of the science was desirable to support the future management of this important fisheries resource. The Habitat Conservation Office requested that the Center for Independent Experts (CIE) of University of Miami conduct a review for the NOAA Chesapeake Bay Office's Blue Crab Stock Assessment.

### Review Activities

The review workshop was conducted at the Radisson Hotel in Annapolis, Maryland over the three days: Tuesday 9<sup>th</sup> August to Thursday 11<sup>th</sup> August 2005. Derek Orner of the NOAA Chesapeake Bay Office provided a hard copy of the 1997 and 2005 assessments plus a CD containing an electronic copy of the 2005 assessment plus a multitude of associated relevant documents and data files, including the model and its code.

The review team and the principle scientific team contributing to the review included:

The review panel, made up of Malcolm Haddon from the University of Tasmania, Paul Medley, a private fisheries consultant from the U.K., Nick Caputi from Western Australia Fisheries, and Mike Bell, a private fisheries consultant also from the U.K. Useful and clear presentations relating to the data and new assessment were made by Thomas Miller, of the Chesapeake Biological Laboratory, with additional discussion and contributions from Lynn Fegley and Glenn Davis of the Maryland Fisheries Administration.

The review process consisted of an introductory day consisting of a presentation of the completed 2005 assessment, with general questions. This was followed by a more intense day involving continuous questions and answers concerning details of the assessment document and related work conducted as four sessions during the day with short breaks in the morning, at lunch-time, and in the afternoon. The final day was devoted to the review panel discussing the review, the presentations, and documents. While there were no further presentations there were some additional, relatively informal discussions especially with Derek Orner and his colleagues from NOAA when they kindly arranged for the review panel to see Chesapeake Bay at first hand, onboard a NOAA research boat. Blue crab fishing was much in evidence during this short tour.

. The timetable of presentations and discussions was:

**Tuesday 9<sup>th</sup> August 2005**

Welcome and Introductions

Derek Orner

Presentation of 2005 blue crab assessment

Thomas Miller

**Wednesday 10<sup>th</sup> August 2005**

Terms of Reference Review and Discussion

Thomas Miller  
Lynne Fegley  
Glenn Davis

Detailed discussion of:

Assess and quantify the life history and vital rates of blue crab in Chesapeake Bay

Fishery-Independent Surveys

Patterns in Catch and Effort by sector and region

The Assessment Model for the Chesapeake Bay Blue crab fisheries.

Evaluation of the control rules for the Chesapeake Bay blue crab fishery.

**Thursday 11<sup>th</sup> August 2005**

Review panel to discuss assessment and methodologies and initiate development of individual review documents.

The format of the review explicitly followed the terms of reference and followed four main themes, with particular emphasis being placed on the three separate aspects of the data used in the assessments:

- 1) Evaluate the adequacy and appropriateness of all data used in the assessment, including the following:
  - Life history and vital rates of blue crab in Chesapeake Bay.
  - Patterns in fishery-independent surveys.
  - Patterns in catch and effort by sector and region.
- 2) Evaluate the adequacy, appropriateness, and application of the assessment models used for the Chesapeake Bay blue crab fisheries and characterize the uncertainty in the assessment.
- 3) Evaluate the scientific basis for the control rule for the Chesapeake Bay blue crab fishery.
- 4) Develop recommendations for future research for improving data collection and the Chesapeake Bay blue crab assessment.

All reviewers considered all aspects of the review but it was agreed that especial emphasis might be expected in the particular areas of expertise for each of the four

reviewers. The issues and details involved with the Chesapeake Bay blue crab assessment meant that meeting the scientists involved in the research that contributed to the assessment was of great value in understanding the content and implications of the assessment. A simple review of only documentary material relating to the assessment would have been insufficient to provide a sufficient grasp of the issues and their present status.

No new sensitivity runs for the assessment were requested by the review panel during the meeting, but each panel member was able to explore the assessment model and data using the files provided on CD.

The review process was conducted in an open and friendly atmosphere with great interest and enthusiasm being expressed by the assessment team in Maryland. The Chesapeake Bay blue crab assessment is a multi-faceted and difficult problem and great progress has been made since the first assessment in 1997 (Rugolo *et al.*, 1997). In line with the tradition of excellent work conducted in Chesapeake Bay, the enthusiasm and openness to critical discussion does both the current stock assessment team (Miller *et al.*, 2005) and their respective organizations credit.

We would all like to thank the contributors and others at the review meetings for their efforts and friendly openness during this review. The provision of documentary materials was also very thorough and appreciated.

#### **DISCLAIMER**

The information in this review has been provided by way of review only. The authors make no representation, express or implied, as to the accuracy of the information and accepts no liability whatsoever for either its use or any reliance placed on it.

## **Summary of Findings**

### **Structure of Document**

Each member of the review team produced their own independent document detailing their individual reviews of the 2005 assessment document and associated material. This present document represents a synthesis of the individual panellist's review reports. In accord with the Statement of Work Annex 1 (see Appendix 2, this document) there will be no attempt at reaching or developing a consensus on the 2005 assessment but rather this document will attempt to fairly summarize the individual Review Reports and focus on collective conclusions and recommendations. No attempt will be made to ascribe any particular opinion to a particular reviewer so this document must be viewed as a synthesis from the four separate individual reviews. Akin to the individual reviews this synthesis will be structured to parallel, approximately, the details of the terms of reference for the review.

There will thus be a discussion of each of the data sources, the population and biological properties of blue crabs in Chesapeake Bay, the time series of fishery independent survey data, and the different fishery dependent data with respect to catch and effort by sector and region. This will be followed by a detailed discussion concerning the structure and appropriateness of the assessment model, then the control rules that stem from the assessment process. And finally, there is also a discussion concerning recommendations for future research and how data collections might be improved. The exact terms of reference are included under the Objectives of the CIE Review described in the Statement of Work (see Appendix 1).

### **Terms of Reference**

**Evaluate the adequacy and appropriateness of all data used in the assessment, including the following:**

#### **Life History and Vital Rates of Blue Crab in Chesapeake Bay.**

A discussion was had involving all details of major relevance to the stock assessment modelling. These included:

- the stock structure of blue crabs within Chesapeake Bay and beyond,
- the revision of the natural mortality estimates,
- the description of blue crab growth,
- the movement and reproductive migrations of the crabs,
- the life cycle and behaviour of blue crabs.

#### **Stock Structure**

That the Chesapeake Bay population of blue crabs can be treated as a predominantly independent stock was supported by the available genetic evidence, the evidence concerning larval dispersal, and by the known movement patterns of adults being restricted to within the Bay and tributaries. It is important to establish that a fishery is based on a distinct stock otherwise inferences made about productivity (growth, recruitment, and natural mortality) may be confused and prone to error because of immigration from outside the study area.

## Natural Mortality

A significant improvement of the 2005 assessment over the 1997 assessment stems from the attention paid to the estimation of natural mortality; a fundamentally important parameter for any fish stock. The importance of the revised estimate of natural mortality to the latest assessment is reflected in the time devoted to this issue both in the stock assessment document (Miller *et al.* 2005) and during the review. The array of methodologies applied to this problem included the use of mark-recapture experiments, the use of lipofuscin, observations on captive animals in experimental ponds, and various approximations based on other life history characteristics such as life span, growth rates, and age at maturity (Hewitt & Hoenig, 2005). The weight of evidence pointed to a marked increase over the assumed mortality rate used in the 1997 assessment and this has had an equally marked effect on the assessment. The decision to bracket the values used in the assessment between 0.6, 0.9, and 1.2 was a sensible compromise considering the uncertainty and differences exhibited between the array of different estimates. The preferred range of natural mortality values lie within the limits from all methods, but are mainly supported by a Delaware study and tagging studies of mature females. The arguments put forward against the natural mortality being a relatively small number (e.g., 0.375 as in 1997) were convincing, especially as the low estimate used in 1997 was based at least in part on limited and suspect tagging data.

Classical tagging studies are only likely to provide good information about life history characteristics if catches can be more closely monitored. Coded wire tags would be the ideal option, but would require a significant proportion of the catch to be monitored for tag returns. This might be done if there are particular processing facilities that handle a significant proportion of the total catch.

The assumption of a natural mortality rate of 0.9 implies that about 60% of each age-class dies naturally each year. This assumption is implemented in the model as a constant rate across all size classes; a common assumption in most stock assessment models. An explicit discussion on this issue was a part of the review meeting. It was concluded that an obvious area for future research would be in size-related natural mortality rates. Intuitively, it seems likely that smaller crabs suffer greater natural mortality than larger crabs. It is known that cannibalism is a greater threat to small crabs than to larger ones. Predation pressure and other sources of natural mortality are likely to be very different for juvenile blue crabs in sea grass and other complex habitat types than for adults in more open environments. Similarly, spawning females undertaking large-scale movements are likely to experience a different mortality regime to adult males. This would be reason enough to explore size-related and gender related mortality possibilities, even if only theoretically by exploring the implications of plausible scenarios in the assessment model. Preliminary trials of such sensitivity tests indicate the assessment is extremely sensitive to size-related differences in natural mortality. This in itself may suggest alternative approaches to the stock assessment modelling. At present the total mortality appears to be so high that very few older crabs survive in significant numbers so these sensitivity analyses may not lead to great insights. It would also be extremely valuable to know whether there is any density dependent release from cannibalism (a major source of natural mortality) when stocks are relatively low. Such a density dependence on natural mortality would provide a compensatory increase in productivity when stocks were low, which would mean the stocks were more robust to exploitation than without such a response.

Overall there have been significant advances in the treatment and estimation of natural mortality since 1997. The relatively high values now used imply that the fishery will have to be very intense to have a detectable effect. This may lead eventually to a need for an assessment model having some form of intra-annual time-step rather than the annual time-step currently used.

## **Growth**

A major contributor to a stock's productivity is the growth of animals already recruited. The description of growth in the latest assessment is far more comprehensive than that used in the 1997 assessment. Not only were alternative approaches to describing growth investigated (using approximations that assumed continuous growth as well as explicitly modelling the moulting process; Bunnell & Miller, *in press*) but a review of a multitude of different growth studies was undertaken. This latter was relatively uncritical but was used primarily to define the possible range of values of growth for the younger age classes 0+, 1+ and 2+. Emphasis was rightly given to studies where growth parameters were estimated directly (length-frequency analysis, lipofuscin calibration of age) rather than those depending on external assumptions. However, although it provides an interesting and innovative way to measure age, the lipofuscin growth estimates are probably relatively inaccurate compared to other methods because lipofuscin can build up in tissues for reasons other than age. One of the fastest growth models included was undertaken in a pond mesocosm (Ju *et al.*, 2001), which may not be representative of growth in the wild (so whether it should be included as a viable possibility is debatable).

Continuous growth functions provide a poor approximation to discontinuous crustacean growth patterns over short time scales. For this reason a moult process model was developed for use in calculation of biological reference points. The moult process model did not depend on the results of recent growth studies, but was partially validated by comparison of observed and predicted size distributions. An important component of the model is the accumulation of degree-days required for moulting. A fuller description of the moult process model is given in Bunnell & Miller (*In Press*), but the temperature-dependent function is not explicitly given.

Tagging is probably the best alternative available for improving the understanding of how growth may vary around the Chesapeake Bay. Internal coded wire tags are currently being used on hatchery-released crabs, but recapture presents the greatest problem. To be cost effective a significant proportion of the landings would need to be monitored. For estimating growth only relatively few returns are needed, but they should cover a wide number of periods of release.

Overall, the description of growth used in the current modelling indicated slightly lower productivity for the Chesapeake stock than in the 1997 assessment, which would be precautionary. The growth of males and females do not appear to differ significantly up to an age of about 20 months, and with total mortality now being so high, assuming that it is reasonable that growth was the same for males and females. Should the stock ever recover to a point where significant numbers of animals live beyond 20 months, it will become necessary to consider the consequences of sex-related differences in growth (which would affect the per-recruit modelling as well as the assessment model).

There remain questions in relation to growth, especially with respect to the transition of animals from under-sized to legally sized animals and the within season timing of such moulting events. In addition, as with the natural mortality, it would be worth knowing whether there were any density dependent changes to growth rates. Such changes would

significantly alter the stock productivity and it is this that gives this research direction such significance. Given more detail on the seasonality of growth and any density dependency would open up opportunities for constructing a within-year model of the stock dynamics. Given the high rate of total mortality, adding more temporal detail to the model appears to be a strategy that may be worth further investigation.

It is worth noting that the 2005 assessment did not attempt to update or unify previously established spatial, temporal and size criteria in the definition of survey indices. The current diversity of criteria for the different age-classes may accurately reflect the diversity in growth patterns around the Chesapeake Bay but this should be explicitly acknowledged in a re-examination of these criteria given the light of new perspectives on growth and spatial population dynamics.

Once again, there have been significant improvements in the estimation and utilization of growth descriptions in the latest assessment.

### **Recruitment Dynamics**

The other major source of productivity for the stock relates to the recruitment dynamics. The knowledge of the life cycle is thorough and detailed. Knowledge of the movement dynamics of the crabs through the year should aid in the monitoring of the supply of recruits. The time series of megalopae from the Chesapeake Bay Program's monthly zooplankton monitoring program, conducted since 1985, was not directly used in the 2005 stock assessment but may provide an indicator of recruitment. Lipcius and Stockhausen (2002) indicated that the poor recruitment in 1991 resulted in the initial decline in spawning stock in 1992 and that subsequent poor recruitment may have been driven by the low spawning stock. Based on that assessment, a Biological Reference Point (BRP) of the spawning stock of about 3 crabs per tow could be considered as a potential limit reference point. Further work on the stock recruitment relationship should be undertaken. For example, the spawning stock from the VIMS survey could be related to the recruitment from the WDS, which seems to provide a robust indicator of recruitment that is subsequently reflected in the variation in catch. In addition, it is important to assess the impact of environmental conditions on the recruitment as Lipcius and Stockhausen (2002) indicated that the decline in the 1991 recruitment was due to environmental conditions (inferred because the spawning stock appeared to be in good condition at the time). The impact of environmental conditions in the below-average recruitment of the last decade also needs to be assessed. The planktonic survey estimates of zoeal abundance are relatively noisy and, so far, no relation between zoeal abundance and 0+ abundance has been found. It seems possible that the estimates of 0+ abundance from the winter dredge surveys will be a more reliable indicator of recruitment and future yield than the continued study of crab larvae; these two indices should be explicitly compared.

Recent information was not available on fecundity patterns. The results of an earlier study showed only a weak linear relationship between egg numbers and female size (Prager *et al.*, 1990), possibly because of within-season variability. Concern was expressed that density-dependent changes in fecundity may make the 1990 results of doubtful relevance to the current Bay population which is now at a much lower level of abundance. It will be important to develop an improved understanding of the biological and environmental factors influencing egg production in Chesapeake Bay blue crabs, particularly if reference points based on spawning potential are to be applied in stock management. Up to a point, reference points based on spawning potential or biomass should not be too sensitive to absolute fecundity estimates, since they are calculated on

a relative scale. However, departure from a linear relationship with size (e.g., a cubic or other power of carapace width) may have consequences for the individual-based per-recruit model that may be worth further investigation.

The VIMS survey provides an index of mature females on the spawning grounds from 1988 to 2004 that shows a significant decline (Lipcius and Stockhausen, 2002). This time series should continue to be monitored closely and holds the promise of being considered valuable as one of the key BRPs for this fishery.

It is also possible that the knowledge of the movement patterns could be used to improve the interpretation of the various fishery independent surveys. However, the sheer scale of the Winter Dredge Survey (WDS) means that information on movement is likely to come from the survey rather than knowledge of movement informing that particular survey.

### **Patterns in Fishery-Independent Surveys.**

Without the fishery independent surveys the assessment of blue crabs could only be relatively weak. There are four time series of data to use and these vary in spatial scale from a restricted area around Calvert Cliffs, Maryland (Calvert Cliffs Pot Survey) to the whole of Chesapeake Bay (Winter Dredge Survey), and in length of time-series from 16 years (Winter Dredge Survey) to 49 years (VIMS Trawl Survey). Descriptions included in the modelling included:

- The Virginia Institute of Marine Sciences (VIMS) trawl survey (1968 – 2003) restricted to Virginian waters but useful in covering the main spawning grounds. This survey started in 1954 but only data from 1968 onwards have been used because of significant changes in scope and gear.
- The Maryland trawl survey (1977 – 2003; missing 1988) restricted to the Bay edges in Maryland waters.
- The Calvert Cliffs pot or trap survey (1968 – 2003), restricted to one set of narrows in the main channel in Maryland waters.
- The Winter Dredge Survey (WDS) (1989-2003), 1500 stations spread through the entire Chesapeake Bay.

Each survey provides a different spatial and seasonal perspective on the population dynamics of blue crab in the Bay, which is reflected in different aggregations of months and areas and different size criteria for distinguishing age-classes (Table 4 of the 2005 Assessment) in providing abundance indices.

The WDS is especially impressive through its sheer scale (about 1,500 stations throughout the Chesapeake Bay), which provides it with very great statistical power to discriminate stock sizes between years. In addition, numerous depletion experiments are conducted in different parts of the Bay each year in order to estimate the relative catchability of the different gears and vessels involved in the surveys (Volstad et al. 2000). Despite these latter efforts it is still surprising that this survey can be used as an absolute estimate of abundance for the 0+ and 1+ age classes. The WDS in December-March has proved able to predict the catch from April to December with an  $R^2$  of 0.85. This indicates that the survey is providing a reliable indicator of abundance. The strength of this relationship also indicates that if there has been any variation in effort over the last decade then this has not had a significant impact on catch. If trends in nominal effort can be obtained for this period it will provide some insights on the

impact of effort on the catch taken. For example, if there has been a significant increase in effort with little impact on catch then it may be an indicator of gear saturation. It would also be useful to investigate the recruitment-catch relationship at different spatial scales such as the State level. This would enable each State to suggest to commercial and recreational fishers the likely catch in the coming season. These relationships may also provide some insight into spatial trends and migration.

Beyond the clear value in providing an abundance estimate upon which to base the assessment, the winter dredge survey has the potential to provide useful insights into the stock recruitment dynamics, depletion rates of the 0+ animals (which obviously become 1+ in the subsequent year), plus an early warning of spatial depletion of particular areas within the Bay. For example, this survey identified that the central mainstem of the Bay has experienced a strong decline for the last decade. This is an indication of the reduction in the spawning stock. While the WDS must be expensive to run, it is recommended that it continue because it provides the best quality and most spatially extensive data set for the assessment of the Chesapeake Bay blue crab stocks available.

Some discussion was spent on trying to suggest ways of improving the catchability experiments. This included the suggestion of using video observations to see if it would be possible to independently estimate the number of animals in the path of the dredge. It seems likely that video observations would only be possible in a few places in the Bay as underwater visibility is usually poor. In addition, the use of differential GPS and GPS plotters should be used instead of Dann buoys to mark the ends of the dredge runs that are to be depleted. While estimating catchability for the gear can help correct estimates, estimating catchability will always be a significant source of error where it varies considerably from survey to survey. Where possible, surveys should be conducted with the same vessels and gear at the same time each year. Minimizing the variation in covariates will minimize errors in applying corrections.

This, in combination with standardized data forms, plus recording details of depth fished, sea-state, and rope-length to the dredge would permit the use of a Generalized Linear Modelling framework to standardize the analysis and provide catchability estimates less affected by spurious factors. The GLM can be simultaneously fitted to the depletion experiments used to calibrate the survey as well as the survey data itself. However, an important part of applying corrections is to build an appropriate set of covariates that can be linked to the survey data.

Covariates should include habitat type, weather conditions, as well as the gear and vessel characteristics. These covariates could be characterized retrospectively, assuming benthic habitat (e.g. particle size, depth, salinity) has not changed over time and standard time series indices such as wind speed and water temperature can be used to represent general weather conditions in which the dredge operates.

The depletion experiments used to calibrate the survey could also be used to estimate selectivity. Unfortunately sample sizes can become small when the catch is broken down into categories. A generalized linear modelling approach becomes more important in these circumstances. This would allow a more sophisticated modelling approach to the survey and should yield more information about the stock.

If the catchability experiments can be improved, the confidence in the survey abundance estimates would be improved. The uses to which the data from this survey is put will be further discussed when the assessment model is considered. Whichever estimation method is used for catchability, it is recommended that the variance of the catchability

estimates be accounted for in confidence intervals presented for the absolute abundance estimates, and that the full uncertainty in abundance estimates be carried forward into other computations such as the calculation of exploitation fractions.

The remaining fishery-independent surveys have been run for longer periods but in each case cover a smaller geographical area and are not so precise in their estimates.

The Maryland Department of Natural Resources Trawl Survey has varied its spatial coverage from year to year (with the timing of the survey within years also varying occasionally). These variations imply this time-series is a less reliable and noisy index of crab abundance through time. Again, it would be possible to conduct a Generalized Linear Modelling exercise to produce an annual index of relative abundance rather than continue to use the ad-hoc averaging across the various rivers and bays surveyed each month. This may improve the analysis produced and would certainly make it more defensible.

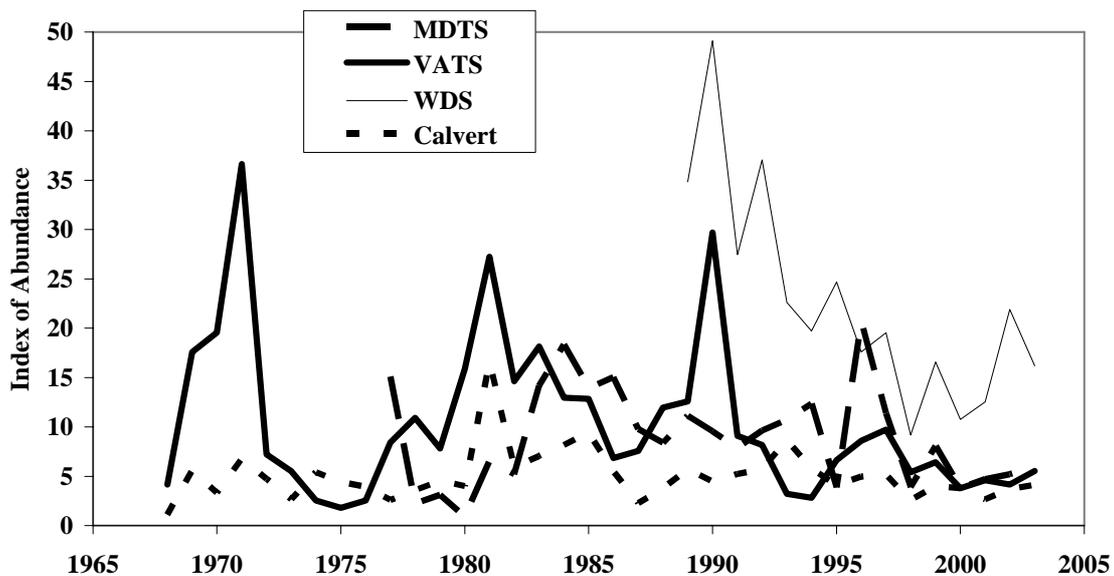
The VIMS survey uses a standard bottom trawl survey and has the longest time series. The spatial scope has expanded, as it started with the main rivers on the western shore and is now much enlarged. Gear changes also occurred through time, and include the addition of a tickler chain in the 1970s, but most of the changes to gear occurred prior to 1968. The gear changes were not made without exploring how the changes affected fishing performance and corrections have been made to the time series of data as a result of these explorations. In particular, all the pre-1968 changes to gear led to the recommendation to exclude all data from pre-1968 from the assessment. To make the data adjustments the VIMS staff conducted paired tows to make the comparisons, but a full GLM analysis was not used and this could have improved the corrections.

The effectiveness of the VIMS survey for estimating crab abundance has been questioned because it was not designed purely to study crabs but is rather a juvenile finfish trawl survey and crabs are only a bycatch. Indeed, the estimated abundance of 0+ is invariably less than that for 1+ crabs, indicating a difference of either availability or catchability. Nevertheless, the estimates for the 1+ crabs can probably be used validly and as such provides a valuable index of mature females on the spawning grounds from 1988 to 2004 (showing a significant decline; Lipcius and Stockhausen, 2002). There is, therefore, still some work required on how best to develop a better indices of relative abundance for both the 0+ and 1+ crabs.

The final fishery independent survey has been conducted at the Calvert Cliffs using pots to follow the relative abundance of crabs through time. This survey began in effort to monitor the potential environmental impact of building a power station that expelled warmed water into the Bay. The trap or potting method relies on crab movement and it may be affected by other pots set locally. Nevertheless, this survey has a long time series and even though it is extremely restricted geographically, there appears to be some relationship between the trends seen in the Calvert Cliff crab pots and the Maryland trawl survey so there may be valuable information in the data with respect to early trends in crab abundance.

The trawls used in the Virginia and Maryland trawl surveys differ, and both obviously differ from the pots used in the Calvert Cliffs and the dredges used in the Winter Dredge Survey, so it is fair to ask whether their results are comparable (Fig. 1). When the different time series are compared there are clear consistencies as well as clear inconsistencies. Some of the features observed in the time series are especially influential in the stock assessment modelling. For example, much of the apparent ability

of the stock to recover from depletion derives from the spike in the abundance index found in the Virginia trawl survey in 1970 and 1971. Because of the issues of comparability and variation between surveys, a formal exploration of the extent of noise and or bias, and the overall reliability of the different surveys should be produced. At very least a more complete set of sensitivity analyses should be conducted to determine the relative contribution of the different data sets. For example, when the Virginia trawl survey results are excluded from the assessment, it was reported that there is a large effect because of the influence of the early years. It might be worth exploring what would happen if one were to remove the spike in apparent abundance derived from the Virginia trawl survey in the early 1970s, as a sensitivity analysis it would be a reasonable idea to include the VATS but exclude the earlier years



**Figure 1.** A visual comparison of the four time series of abundance indices for blue crabs in Chesapeake Bay. Some consistency is visible but also some inconsistency. The large spike of productivity indicated by the Virginia trawl survey around 1970 is very influential in the modelling.

Further analysis of the coherence between surveys is required. It is the case that each survey places different emphasis on different aspects the same thing. For example, the Virginia trawl survey measures more females than the Maryland trawl survey simply because of timing of the surveys combined with the spatial distribution and migration patterns of the two sexes. So are they measuring the same thing? Does each survey provide a different perception of the stock that would include implicitly the spatial details of the stock? The GIS analysis of the WDS suggests a spatial structure to abundance that should not be ignored (Jensen & Miller, *In Press*; Jensen *et al.* (a) *In Press*; Jensen *et al.* (b) *In Press*). But perhaps the annual time step addresses this in a valid fashion. Certainly if a seasonally explicit stock assessment model is ever developed, then spatial details for the different surveys could be used separately.

It is recommended that the WDS be given more weight in the assessment modelling because of its extensive coverage and precision. However, in all comparisons there needs to be at least some overlap between the time series being compared. For example the WDS exhibits a strong decline over the last 10 years and this is reflected in the other series (although to a lesser extent).

Currently the different time series are being treated as if they were effectively independent. It would be worth formally exploring relationships and correlations between the different series. Because of the different quality of the surveys it is suggested that the WDS be kept separate from the remaining surveys. Some formal statistical method should be used to combine the three remaining time series (perhaps a straight-forward generalized linear model or perhaps some multi-variate ordination method).

One possible way of weighting the different survey time series would be to examine the relationships between the various surveys and subsequent yields (possibly of particular age classes. For example, there is a demonstrated strong relationship between the WDS and subsequent yield. It would be interesting to determine whether, for example, the Virginia trawl survey is related to the subsequent yield of female crabs.

The survey data are used in the assessment by converting each to a series of standardized Z-scores, which assumes they vary about their mean values in an approximately normal and symmetrical fashion. This appears to be inconsistent with using a log-normal transformation in the stock assessment modelling.

Finally, in addition to the fishery independent survey time series there are also numerous time-series of sea-water temperature. These may have value for the exploration of possible relationships between relative recruitment levels each year and environment variation. This could be examined through a consideration of both the larval survey and the estimates of 0+ crabs especially from the WDS.

#### **Patterns in Catch and Effort by Sector and Region.**

There are now regulations requiring more accurate details concerning catch and effort, at least in the commercial fishers. There have been detailed efforts (Fogarty & Miller, 2004) made to understand the implications of changes to regulations and the results are an improvement over the previous situation. Despite the new regulations it is known that there is a degree of under-reporting of catch, especially in the peeler fishery (in which any mortality during the shedding process tends to get omitted). There is a myth/tradition that peeler crabs are far less effectively reported than the hard crab. How true this myth is in practice should be determined.

A cross check on the accuracy of the catch estimates is being examined in Maryland using dealers catch reporting and voluntary daily reporting of catch rates by selected cooperating fishers and observers on board commercial vessels. This work should continue to be encouraged and extended into Virginia.

While it could be extremely useful to obtain accurate catch rate and effort data it is essential to have accurate estimates of the total extractions from the fishery. The continued lack of accurate catch and effort information from both the commercial and recreational fishers makes management of total effort full of uncertainty.

The problem of obtaining accurate catch and effort statistics are exacerbated by the different spatial distributions of the male and female crabs leading to different degrees of fishing intensity on the two sexes. Catches of hard crab are now reported by sex except in the peeler fishery (which is likely to be biased towards females). Overall, the selectivity by sex was reported as being approximately 50:40:10 for female:male:peeler. It would appear that more work is required to obtain better information from the peeler fisher (for both total landing and the sex ratio of the catch).

Estimation of the Bay-wide catch in numbers involves a series of data aggregations and conversions between weights and numbers. Each stage of the process involves its own uncertainties, and it would be useful to see how these translate into overall uncertainties for the Bay-wide catch numbers and uncertainties in the exploitation fractions. An important assumption underlying the conversion of hard crab weights to numbers is that annual, sex-specific mean carapace widths from the fishery-independent trawl surveys are representative of the sizes of crabs in the landings. This is unlikely to be an important source of bias in the calculation of catch numbers, but it would be worth validating this assumption using measurements of crabs sampled from the commercial landings.

In 2002 there were reported to be 6170 licensed crabbers. Unfortunately, about 13% of these provided no catch records, while about 47% explicitly reported zero catches. Out of the total, about 17% reported less than less than 20 bushels in the year while about 24% reported more than 20 bushels. There remains an issue of measurement units where bushels are often converted first to pounds and then to numbers. The potential for making errors appears to be relatively high so it would be useful to determine whether the estimates of catch are invariably biased low (as would appear to be likely).

Currently there are estimated to be between 1,400 and 1,500 active fishers in Maryland with similar numbers in Virginia. In May 1999 a cap was put of the entry of new licenses but new entries may simply be using up latent effort (making inactive licenses active). It is unfortunate that the regulations in the fishery differ between Maryland and Virginia, because this must make obtaining comparable data from the fishery difficult. For example, in Maryland ovigerous crabs (so-called 'sponge' crabs) may not be landed while in Virginia the regulation forbids only dark sponges (which would be close to spawning). Another example is that escape gaps (cull rings) are required in Maryland (though they need not be open!) whereas they are not required in Virginia. A review of regulations with an aim of increasing commonality between the two States may improve the ability to collect consistent data.

Apparently there is now a new problem of some fishers using excess numbers of pots by attaching multiple pots to single buoys. This source of extra effort, as well as being illegal, would be very hard to detect with current data collection methods.

An observer program, started in 2002, is used to verify catch rates and could be used to corroborate total catches. In fact, this program indicates that reported catch rates are very close to the actual catch rates. The observers have been able to demonstrate some regional differences in error rates for catch rates (some of the more isolated island fishers exhibit a large bias downwards in their records). The catching characteristics of peeler pots and hard crab pots have been found to be very different.

Catch has also been reported from dealers since 2002. For hard crabs reported landings recorded by dealers were close (within about 10%) to reported catches from fishers. However, there are known biases, for example there is a direct basket trade (direct sales from boats to consumer). In addition, females are often processed for meat but the males (a preferred product) are somewhat under-reported because of the direct sales.

Estimating the total catch is important so that the dynamics of the stock can be more precisely modelled and described. Errors and bias in the estimates of total catch could constitute a serious under-estimate of total extractions so the impact of the fishery on the stock may also be under-estimated. Related to under-estimated total catches are other forms of mortality associated with the fishery. Discarding of undersized crabs is

common and whether gear associated mortality of the under-sized crabs is significant is unknown. This would be worth exploring. Another potential source of unaccounted for mortality could be a result of ghost fishing by lost crab pots. Apparently there are significant numbers of pots lost each year. If lost pots are self baiting (e.g. fish entering the pot and then dying) then this could constitute another significant source of unaccounted mortality. It is strongly recommended that the extent of ghost fishing be assessed with regard to the potential losses and how, if they are significant, they may be ameliorated.

A potentially controversial issue related to unaccounted mortality relates to the wide array of different fishing methods used to catch crabs. If some fishing methods are harder on the crabs than other methods then it would be sensible for an examination of the potential relative impact of those different gears in terms of unaccounted mortality. Such a study may be valuable in estimating the total impact of the different gears used once accurate estimates of effort by gear become available.

Recreational catch was not included in the stock assessment as it was not available annually. It was estimated for 2001 and 2002 as being about 5-9% of the total harvest. A sensitivity analysis was undertaken to assess the impact of this additional harvest. While there have been recreational surveys, it is not known if there is a relationship between stock size and recreational take. (In other words, does recreational take increase with stock size in line with commercial catches?) A second survey is being undertaken and this may provide some enlightenment with regard to this process. There is a recreational license for trot-lining fishers. The people who buy the licence tend to be really serious with their fishing. Thus, while the bulk of the effort comes from the mass of the public, the bulk of the catch comes from those with licences. It might be useful to monitor the licences sold to trot-liners each year as an indicator of the trend in recreational effort. The catch and effort trend of this group may make a significant proportion of the total recreation catch. A random sample from this group could also be selected to undertake a phone/diary survey of their fishing activities during the year. The proportion of total recreational catch this group takes should be available from the surveys of the whole recreational fishery.

The important thing is to be confident in the coherence of the different time series of catch estimates being developed. This coherency should be considered explicitly. The difference between the improved catch estimates and the traditional catch estimates should be explored to determine how far normal catch estimates might be from that reported.

The absence of effort data does not detract from the value and validity of the 2005 assessment, but it will be important to address this issue in future. The principal value of effort data would be that it might allow effective management tools to be designed. The most worrying aspect of the catch and effort data is the impression that effort has expanded recently to unprecedented levels. If this really is the case, then methods obviously need to be developed to reduce effort to more manageable levels.

### **Evaluate the adequacy, appropriateness, and application of the assessment models used for the Chesapeake Bay blue crab fisheries and characterize the uncertainty in the assessment.**

The stock assessment model used to assess the Chesapeake Bay blue crab is a modified version of the Collie & Sissenwine (1983) model. This is a simple model structure using a two-stage population structure (pre-recruit and fully-recruited), which was originally

developed to be used with fisheries for which only survey information is available. Applications of this model include assessment of American lobsters (Conser & Idoine, 1992) and northern shrimp (Cadrin *et al.*, 1999). Recently the method has been advocated as an alternative to fully age-based methods for species in which age-determination is difficult or age composition data are missing (Mesnil, 2002, 2003a). In the 2005 blue crab assessment the model has been modified to include the array of different fishery independent abundance indices derived from the various surveys conducted in the Bay. The model has been implemented in AD-Model Builder (Fournier, 2000) and the application code and data files were provided for the review. This use of multiple time-series of data is another valuable innovation and is the reason the model is referred to as the Catch-Multiple-Survey (CMS) model. The implementation uses data relating to catches from 1968 to 2003, using an annual time step. It uses the fishery independent estimates of exploitation rate ( $\mu$ ) from the WDS from 1990 to 2003. In addition, it uses data from both the Maryland and Virginia trawl surveys (but does use the data from the Calvert Cliffs; though the option exists). This is a reasonable use of available data and attempts to include as much of the broad-scale data as possible.

The contributions of the various surveys used in the assessment are weighted by the respective variance of the observation errors within each time series, which is appropriate for this model structure. However, it would be useful to know whether this weighting affected the balance between the observation and process errors in the model.

The CMS model requires estimates of juveniles and adults but does not appear to pay account to the different selectivity of the survey gear for the 0+ and 1+ crabs. This may be a mis-perception however, as the survey estimates may have been made with this in mind.

With regard to the calculation of exploitation fractions from the WDS and estimated catch numbers, it can be re-iterated here that both the fishery and fishery-independent data sets contain uncertainties, and for the purposes of (a) robust inference of stock and fishery trends, and (b) risk assessment in fishery management, it would be useful to include that uncertainty (e.g., as confidence intervals) in presentations of abundance and exploitation estimates. The value of including the other survey indices is that it permits an extension of the analysis of stock trends right back to 1968 (the WDS only extends back to 1990). This may be introducing biases from the various surveys and emphasizes the need to examine the cohesion between the different survey indices. Perhaps the best option would be to combine all time-series (except the WDS) into one (using a Generalized Linear Modelling approach) and scaling that index to that produced from the WDS.

The CMS model includes two major developments over the original CS model. Firstly, instead of estimating stock size separately for each year, the CMS model estimates stock size only for the first year, with stock sizes in subsequent years being derived from the population dynamic model (i.e. the survivors from the previous year's stock and recruitment). This allows use of multiple series of pre-recruit and fully recruited stock indices, each of which may be of different length and have missing values for some years. The second development is the maximum-likelihood estimation of  $q_n$ , the catchability index for fully recruited crabs in each survey series, essentially as a geometric mean over years of the ratio between the fully recruited survey index and the estimated stock size in each year.

Essentially the same model formulation has been derived independently by Benoit Mesnil (Ifremer, Nantes, France), described in an unpublished program documentation presented at an ICES Working Group meeting (Mesnil, 2003b), but was not fully developed for application to multiple survey indices. The Mesnil model application also differs from the CMS application to blue crabs in that it does not use a mixed error approach, minimizing measurement error only. Mesnil (2003b) cites several reasons for excluding process error, including difficulties in defining appropriate error terms and weighting factors. Earlier applications of similar models have also excluded process errors (Collie & Kruse, 1998).

The representation of the process error in the CMS model could be improved. The current attribution of the process errors being random and log-normally distributed appears to be arbitrary although it does have the advantage of avoiding the potential for negative mortality (leading to population growth).

The research team highlighted an inverse relationship between exploitation fraction (Catch/Abundance) and abundance, which is suggestive that the crab fishery behaves in a depensatory manner. This may be a result of the indices fluctuating with greater variance than the catches. The process could occur through the concentration of the stock during periods of lower abundance or if the fishers are better able to target the remaining stock concentrations as the stock declines. As the population varies with the indices, relatively constant catches will be explained by changing the effective catchability. It is also possible that the population model is simply inadequate and these results indicate that a model with density dependent mortality may fit the data much better. If the proposed depensation is real, this pattern could have important consequences for fishery management – as the report points out, such depensation is not conducive to sustainability. It will be important to determine whether the apparent depensation is a real feature of the data, or is in fact an artefact of the method of calculation. There will inevitably be a negative relationship between exploitation fraction and stock size, since the first variable is calculated using the second variable as a denominator. This may or may not be enough to account for the relationships shown in the control plots of Figs 39-42 of the 2005 assessment. A simple randomization test may be enough to determine whether the observed negative correlations are greater than would be expected by chance given the nature of the calculation.

Fishing is unlikely to have a measurable impact in the annual model as mortality rates are too high for a significant proportion of the stock to survive from year to year; making an annual assessment suspect. On the other hand, mortality rates are high enough to detect within-season depletion. Within-year timings of indices are probably important to management of the fishery. An annual model can only cope with this approximately, but a within-season model would be able to account for seasonal movement, mortality and growth explicitly. This is certainly a direction that could be explored in the future.

Differences between male and female stock dynamics are not included in the assessment model. Male and female blue crabs show different patterns of movement within Chesapeake Bay, have different patterns of moulting and growth after maturity, and presumably experience different regimes of both fishing and natural mortality. Exploitation fractions are substantially higher for females than males (Fig. 37 of 2005 Assessment). The aggregation of data across males and females may or may not cause problems for the assessment, and there may well not be sufficient data for adequate parameterisation of assessment models for the sexes separately. Nevertheless, it is

recommended that the potential impacts of male-female differences on the assessment outcomes be examined further.

An obvious improvement would be to see some estimates of uncertainty around assessment outcomes, e.g., confidence intervals around the stock abundance estimates. Bootstrapping would be a suitable method to estimate both confidence intervals and bias in the estimates. Ideally, this uncertainty would be carried forward into comparison of assessment outcomes with biological reference points – i.e. the context of risk assessment for fishery management. While the current assessment is adequate for the purposes it is being put to, it may be possible in the future to develop a Bayesian interpretation of the model, which would open up the possibility of explicitly including the uncertainties in the estimates of parameters such as natural mortality and growth. However, investigation of the importance of uncertainties on the model outcomes should be explored before contemplating such modifications.

In all there are 206 data points used in the model fitting process. The reason the number of data points is important is that there are 76 different parameters being fitted in the full model. This is not a particularly high density of data points for the number of parameters being estimated (about 2.71 data points per parameter). The parameters are the initial population size at the start of the modelled period, three parameters relating to the ratio of measurement to process errors, 37 estimates of the abundance of 0+ animals and 35 estimates of a component of the process error in each year; a total of 76 parameters.

The model is implemented using both measurement or observation errors and process errors. The interpretation of the meaning of the process errors is complex but perhaps the best interpretation is that they attempt to capture the variation in natural mortality that occurs from year to year.

It turns out that the model as it stands is not particularly stable. It seems best to run the model using only the adult crabs (1+) in the estimates of exploitation rate ( $\mu$ ) and without including the comparison of predicted versus observed exploitation rates in the likelihood function being used to fit the model to the available data. If both adult (1+) and 0+ crabs are used in the calculation of the exploitation rate, then the model generates negative exploitation rate estimates (not a reasonable outcome). On the other hand, when only the adults are used there are instances of exploitation rates being predicted at levels greater than 1.0 (another impossibility – meaning more than 100% of the available biomass was taken in the fishery). This latter phenomenon seems to imply that the definition of fully recruited crabs is not correct (meaning that some 0+ crabs may be as available to the fishery as some 1+ crabs). This sensitivity to which abundance estimates are used needs to be investigated especially as the 0+ crabs are a major part of the signal in the WDS.

While it is not in the terms of reference for this review to suggest an alternative assessment, it would be valid to point out that some models inherent in the current model could be profitably explored further. The inclusion of process errors in the model adds 35 parameters that require estimation and these appear to de-stabilize the model. In addition, only using observation errors has the advantage that it is possible and stable to include the observed versus predicted exploitation rates in the model fitting. Other studies have found that with simple models it is often best to only use observation errors (Walters & Ludwig, 1981). Mesnil (2003b) has shown it is possible to run such a model without process errors. Appendix 3 of this document also illustrates some trials with variants on the current stock assessment model and these also show that process errors

are not necessary to fit the model. Including process errors has the advantage of comparability with earlier models, but before including such errors and their associated parameters their use should be fully justified.

The documentation of the model needs to be updated to include the full details of the model. For example, in the model code fishing mortality occurs in the middle of the year and the implications of simulating fishing in the middle of the year should therefore be included in the model equations. Within the documentation a detailed justification and need for the inclusion of the process errors should be included. Such a justification should include a comparison of model fits and implications with and without process error. This would provide a far more defensible position and ensure that the balance between useable data and numbers of parameters being fitted is optimized.

Having said all that, the 2005 assessment implements and uses many advances over the 1997 assessment. The 1997 assessment did not use lagged recruitment appropriately, and there has been a re-definition of age-structure size distributions. The implementation of the exploitation rates and the suggestion of using them as a fishery performance measure is a significant improvement over the earlier assessment.

### **Evaluate the scientific basis for the control rule for the Chesapeake Bay blue crab fishery.**

There is a legislative requirement to have in place a set of limit reference points and target reference points against which the fishery may be assessed. The definition implied in Figure 1 of the 2005 assessment document appears robust. Three biological reference points are defined: a) overfishing, a limit reference point, is defined as the exploitation rate at which the spawning potential is reduced to 10% of that in an unexploited stock; b) the target for fishery management is defined as the exploitation rate at which spawning potential of 20% of the unexploited stock is achieved; c) a further limit reference point is the minimum acceptable population size defined as that found in 1968 (acceptable because a fishery was able to continue and recover). The observed abundance in 1968 was a minimum but the model predicts 1975 as having a lower abundance, this year should be considered as a possible alternative limit.

The control rules examined focused on the level of exploitation to assess the level of overfishing and a stock abundance indicator to assess whether the stock was overfished. The minimum population size limit definition depends on a potentially unreliable survey point for 1968. A reference point based on more data points would be desirable. The obvious way to proceed would be to use the residuals from the assessment model, allow for error, and choose indices that would clearly indicate that the stock biomass has fallen below the chosen threshold.

A significant change in the control rule has been the use of the exploitation fraction rather than instantaneous fishing mortality  $F$ , which avoids the need to estimate natural mortality  $M$  to assess the exploitation; although the reference points themselves would need to be changed if a new value of  $M$  was adopted. Because the exploitation estimates are derived from the WDS, and this is the most reliable and accurate survey having the greatest geographical generality, then the control rules are probably as good as they can be at the present time.

To focus on the trends in the exploitation and the abundance and avoid the annual fluctuations, a 3-year moving average of the indicators can be used as part of the control rules. The annual indicators of catch and abundance can be affected by the

environmental factors that can have a positive or negative effect on catchability. Therefore an average over 3 years will avoid the short-term impact of catchability changes and assist in focusing the control rules on the significant trend in the fisheries. At the same time, it would be useful to managers if an indication of uncertainty were to be added to the estimates of the stock abundance statistics. This would enable managers to consider the risk that the true value of the stock abundance was in fact above or below the thresholds.

Currently, the stock assessment model is not used in setting the control rules. The central aim of further development of the stock assessment model would be to provide relevant information for setting the control rule and testing it through simulations.

While the limit and target reference points have been defined, the management actions that might follow if the fishery enters an undesirable state remain unclear. The management levers that can be used to reduce effort appear to be limited to altering the open season (as in Maryland) and through the agency of closed areas (as in Virginia). The effectiveness of the closures in Virginia (Lipcius *et al.*, 2003) has been demonstrated but their effectiveness and that of seasonal closures are confounded with rapid abundance changes. There appear to be limited options for fishery controls but, whatever is used, an explicit decision rule and recovery plan needs to be put in place.

Unfortunately, there appears to be insufficient information on effort and the distribution of effort to improve the management levers or develop new ones. This is a job for the fishery managers not the scientists. The objectives towards which the fishery is to be managed needs to be made clear before suitable means of managing the fishery can be suggested.

In short, the legislative requirements to have limit and target reference points in place has been met but the complementary management levers needed to respond should the reference points be triggered may not be in place. The definition of overfished still appears to be based upon the minimum unstandardized abundance observed in 1968. This data point is quite strongly influenced by the first year of the Calvert Cliffs dataset and the generality of this limit is therefore questionable. What is required is a definition of minimum spawning biomass below which a rebuilding strategy is instigated. The Magnuson-Stevenson Act requires such a recovery plan if such a minimum is defined.

On the other hand, the Chesapeake Bay blue crab fishery may be so variable and unpredictable in the long term that some form of management structure may be impossible unless based upon recruitment processes. The dynamics of short-lived species experiencing naturally high levels of mortality are often defined more by variations in annual recruitment than by variations in mortality (e.g., estuarine bivalves, Van der Meer *et al.*, 2001). The WDS provides a good estimate of 0+ animals. This has been shown to provide an excellent estimate of the eventual yield possible from the fishery in that same year. Rather than some management regime based upon fisheries theory appropriate to a relatively long lived species, it may be more appropriate to examine alternatives that relate more appropriately to short lived, highly variable species. Perhaps a minimum biomass required to generate the harvest could be produced using the strong relationship found between 0+ and subsequent yield. Once again this emphasizes the value and importance of the winter dredge survey and the need to explore the available data for signs of a stock recruitment relationship. If fishing mortality in fact only plays a minor role in the stock dynamics and recruitment variation is more important, then this will have direct effects on management options. The Virginia closures to permit movement and spawning take on greater value if the fishery

is recruitment driven rather than effort driven. If effort levels were better known a plot of catch against effort would exhibit a straight line under conditions of recruitment dominating the dynamics.

Overall, the scientific basis of the control rules for the Chesapeake Bay blue crab fishery is well founded. The switch to using the estimated exploitation rate is a great improvement as it should make it more easily understood by fishers and managers. However, it must be emphasized that no matter how good the control rules, if there are no management actions that can influence (manage) effort or catch should the controls rules indicate the stock is in an undesirable state, then, while the letter of the law may have been met, the intent has been missed. This fishery is in urgent need of some clear and explicit objectives towards which it can be managed.

### **Develop recommendations for future research for improving data collection and the Chesapeake Bay blue crab assessment.**

See Executive Summary

### **Conclusions**

The 2005 assessment constitutes a major advance over the 1997 assessment. There is an impressive array of research projects focussed on blue crab and Chesapeake Bay that have directly contributed to the new assessment. These included:

- (a) natural mortality studies (Hewitt and Hoenig 2005, Lambert *et al.* 2005);
- (b) studies on growth dynamics (*e.g.*, Bunnell and Miller in press, Miller and Smith 2003, Smith and Chang in press);
- (c) a time series analysis of catch that adjusted for changes in reporting methods (Fogarty and Miller 2004);
- (d) use of depletion experiments during the winter dredge survey (WDS) to estimate catchability, which can be used to estimate absolute abundances (Volstad *et al.* 2000, Sharov *et al.* 2003);
- (e) assessment of WDS to quantify abundance and distribution (Jensen *et al.* in press a and b, Jensen and Miller in press);
- (f) reproduction issues such as sperm limitation (Hines *et al.* 2003);
- (g) the Virginia Institute of Marine Science (VIMS) survey provides a valuable index of mature females on the spawning grounds that has been combined with a survey of the megalopae to assess the stock-recruitment relationship (Lipcius and Stockhausen, 2002);
- (h) an assessment of spawning sanctuary and protected deepwater dispersal corridor used to improve the level of spawning stock (Lipcius *et al.* 2003);
- (i) a new stock assessment model that used data from a number of fishery-independent surveys to assess changes in abundance and exploitation;
- (j) individual-based approach to spawning potential per recruit (Bunnell and Miller in press).

These manuscripts and publications have improved the adequacy and the appropriateness of the data used in the assessment. This includes improving details of the biological properties of the blue crabs (natural mortality, growth, and recruitment), as well as the time series of population estimates from the various surveys and the time series of catches that have been amended to account for altered methods of data recording through time.

The assessment model used for the Chesapeake Bay blue crab population is appropriate for a species for which there exists good quality fishery independent survey data of population sizes through time (which matches the blue crab). The modifications to the original Collie and Sissenwine model help match the model to conditions and data available in Chesapeake Bay. They constitute an improvement over the original model. The documentation of the model could be improved slightly and there should be a justification of the use of process errors in addition to observations errors, or the removal of the process errors.

The control rules developed using the exploitation rate or fraction are an improvement over the earlier control rules based on instantaneous fishing mortality (by being simpler to understand). The new control rules are not matched by workable management levers that can be activated should the control rules indicate that the stock is in an undesirable condition. Decisions need to be made (not by the scientists) about how and towards what aims this fishery is to be managed. Some way of limiting the apparently run-away effort currently being expressed is urgently required.

Finally, 20 recommendations are suggested that would include a range of useful directions in which to move future research. The idiosyncrasies of the fishery will help determine which are the most urgent in themselves.

## Bibliography

- Bunnell, D. B., and T. J. Miller. in press. An individual-based modeling approach to per-recruit models: blue crab *Callinectes sapidus* in the Chesapeake Bay. *Canadian Journal of Fisheries and Aquatic Science* (in press).
- Cadrin, S.X., Clark, S.H., Schick, D.F., Armstrong, M.P., McCarron, D. & Smith, B., (1999) Application of catch-survey models to the northern shrimp fishery in the Gulf of Maine. *North American Journal of Fisheries Management* **19**: 551-568.
- Chesapeake Bay Commission (2003) *Blue Crab 2003. Status of the Chesapeake Population and its Fisheries*. Blue Crab Technical Advisory Committee. 20 p.
- Chesapeake Bay Commission (2005) *Blue Crab 2004. Status of the Chesapeake Population and its Fisheries*. Blue Crab Technical Advisory Committee. 8 p.
- Collie, J. S., and M. P. Sissenwine. (1983) Estimating population size from relative abundance data measured with error. *Canadian Journal of Fisheries and Aquatic Science* **40**:1871-1879.
- Conser, R. & Idoine, J., (1992) A modified DeLury model for estimating mortality rates and stock sizes of American lobster populations. Northeast Fisheries Science Center, Research Doc. SAW 14/7, Appendix to CRD-92-07.
- Fogarty, M and T Miller (2004) Impact of a change in reporting systems in the Maryland bluer crab fishery. *Fisheries Research* **68**: 37-43.

- Fournier, D. (2000) *An Introduction to AD Model Builder: For use in Nonlinear Modelling and Statistics*. Otter Research, Sidney, Canada. 178 p.
- Hewitt, D.A. and J.M. Hoenig (2005) Comparison of two approaches for estimating natural mortality based on longevity *Fishery Bulletin (In press)*
- Hines, A. H., P. R. Jivoff, P. J. Bushmann, J. van Montfrans, S. A. Reed, D. L. Wolcott, and T. G. Wolcott. (2003) Evidence for sperm limitation in the blue crab, *Callinectes sapidus*. *Bulletin of Marine Science* **72**:287-310.
- Jensen, O. P., R. Seppelt, and T. J. Miller. (in press) a. Winter distribution of blue crab (*Callinectes sapidus*) in Chesapeake Bay: application of a two-stage generalized additive model (GAM). *Marine Ecology - Progress Series* (in press).
- Jensen, O. P., and T. J. Miller. (in press) b. Geostatistical analysis of blue crab (*Callinectes sapidus*) abundance and winter distribution patterns in Chesapeake Bay. *Transactions of the American Fisheries Society* (in press).
- Jensen, O. P., M. C. Christman, and T. J. Miller. (in press) c. Landscape-based geostatistics: a case study of the distribution of blue crab in Chesapeake Bay. *Environmetrics* (in press).
- Ju, S. J., D. H. Secor, and H. R. Harvey. 2001. Growth rate variability and lipofuscin accumulation rates in the blue crab *Callinectes sapidus*. *Marine Ecology-Progress Series* **224**:197-205.
- Lambert, D. M., Hoenig, J. M., Lipcius, R. N. Hepworth, D. A. and L. Fegley. (2005) Tagging studies of blue crab in Chesapeake Bay: NOAA Chesapeake Bay Office, Annapolis, MD.
- Lipcius, R. N., and W. T. Stockhausen. 2002. Concurrent decline of the spawning stock, recruitment, larval abundance, and size of the blue crab *Callinectes sapidus* in Chesapeake Bay. *Marine Ecology-Progress Series* **226**:45-61.
- Lipcius, R. N., W. T. Stockhausen, R. D. Seitz, and P. J. Geer. 2003. Spatial dynamics and value of a marine protected area and corridor for the blue crab spawning stock in Chesapeake Bay. *Bulletin of Marine Science* **72**:453-469.
- Mesnil B., (2002) Implementation of a two-stage assessment model (Collie-Sissenwine, Conser): presentation of the method and validation of the software. Report of the ICES Working Group on *Nephrops* stocks, Lorient, 3-9 April 2002, ICES CM 2002/ACFM:15, Appendix 2 : 232-242.
- Mesnil, B., (2003a) The Catch-Survey Analysis (CSA) method of fish stock assessment: an evaluation using simulated data. *Fisheries Research*, **63**: 193-212.
- Mesnil, B., (2003b) CSA: Catch Survey Analysis. Assessment Program Documentation. Unpublished report, Ifremer, Nantes.
- Miller, T.J. (2001) Matrix based modelling of blue crab population dynamics with applications to the Chesapeake Bay. *Estuaries* **24**: 535-544.
- Miller, T. J., and S. G. Smith. 2003. Modeling crab growth and population dynamics: Insights from the blue crab conference. *Bulletin of Marine Science* **72**:537-541.
- Miller, T., Martell, S., Bunnell, D., Davis, G., Fegley, L., Sharov, A., Bonzek, C., Hewitt, D., Hoenig, J., and R. Lipcius (2005) *Stock Assessment of Blue Crab in Chesapeake Bay 2005*. Technical Report Series No. TS-487-05 of the University of Maryland Center for Environmental Science.

- Prager, M. H., McConaugh, J.R., Jones, C. M., and P. J. Geer. (1990) Fecundity of blue crab, *Callinectes sapidus*, in Chesapeake Bay: Biological, statistical and management considerations. *Bulletin of Marine Science* **46**:170-179.
- Rugolo, L., Knotts, K., Lange, A., Crecco, V., Terceiro, M., Bonsek, C., Stagg, C., O'Reilly, R. and D. Vaughan (1997) *Stock Assessment of Chesapeake Bay Blue Crab* (*Callinectes Sapidus*). National Oceanic and Atmospheric Administration, Southeast Fisheries Science Center.
- Sharov, A. F., J. H. Volstad, G. R. Davis, B. K. Davis, R. N. Lipcius, and M. M. Montane. 2003. Abundance and exploitation rate of the blue crab (*Callinectes sapidus*) in Chesapeake Bay. *Bulletin of Marine Science* **72**:543-565.
- Smith, S. G., and E. S. Chang. in press. Molting and Growth. Pages Chapter 3 in V. S. Kennedy, editor. *The Blue Crab*.
- Van der Meer, J., Beukema, J.J. and Dekker, R., (2001) Long-term variability in secondary production of an intertidal bivalve population is primarily a matter of recruitment variability. *Journal of Animal Ecology*, **70**: 159-169.
- Volstad, J. H., A. F. Sharov, G. Davis, and B. Davis. 2000. A method for estimating dredge catching efficiency for blue crabs, *Callinectes sapidus*, in Chesapeake Bay. *Fishery Bulletin* **98**:410-420.
- Walters, C.J. and D. Ludwig (1981) Effects of measurement errors on the assessment of stock-recruitment relationships. *Canadian Journal of Fisheries and Aquatic Sciences*, **38**: 704-710.

## **Appendix 1: STATEMENT OF WORK**

### **Consulting Agreement between the University of Miami and Dr. Malcolm Haddon**

*July 21, 2005*

#### **Background**

The blue crab supports the most important commercial fishery in the Chesapeake Bay. Commercial landings have exceeded 100 million pounds historically (1993) with more recent average landings reaching approximately 72 million pounds. The total impact of the blue crab fishery to the Chesapeake region exceeds \$200 million annually.

Sound management of this resource requires accurate information on the status and trends of the blue crab population and on the dynamics of the fisheries that exploit the stock. There have been two recent stock assessments completed for the blue crab (1997, 1998) and the NOAA Chesapeake Bay Office (NCBO) has produced annual 'Advisory Reports' for blue crab to assist resource managers in the decision making process. Seeing the need for an updated assessment, the NCBO supported the development of a full blue crab stock assessment utilizing FY2003 funds.

This assessment was initiated in October 2003. Due to the political nature of any decision regarding fisheries in Chesapeake Bay, especially blue crab, an independent and expert review of the science is necessary for management of this important fisheries resource. The Habitat Conservation Office is requesting that the Center for Independent Experts (CIE) conduct a review for the NOAA Chesapeake Bay Office's Blue Crab Stock

Assessment.

The review workshop for the Chesapeake Bay blue crab assessment will take place in Annapolis, Maryland on August 9-11, 2005. The NOAA Chesapeake Bay Office will provide the following documents prior to the Chesapeake Bay blue crab stock assessment review meeting:

- 2005 Chesapeake Bay blue crab assessment report;
- 1997 and 1998 blue crab stock assessments;
- Annual blue crab advisory reports;
- Adopted management strategies establishing targets and thresholds;
- Chesapeake Bay Fishery Management Plan (1997); and
- Other key publications as necessary.

#### **Objectives of the CIE Review**

The Blue Crab Assessment Review Panel will evaluate the Chesapeake Bay blue crab stock assessment, including input data, assessment methods, and model results. The following are the main terms of reference for the review:

- 1) Evaluate the adequacy and appropriateness of all data used in the assessment, including the following:
  - Life history and vital rates of blue crab in Chesapeake Bay.
  - Patterns in fishery-independent surveys.
  - Patterns in catch and effort by sector and region.

- 2) Evaluate the adequacy, appropriateness, and application of the assessment models used for the Chesapeake Bay blue crab fisheries and characterize the uncertainty in the assessment.
- 3) Evaluate the scientific basis for the control rule for the Chesapeake Bay blue crab fishery.
- 4) Develop recommendations for future research for improving data collection and the Chesapeake Bay blue crab assessment.

The Assessment Review Panel's primary duty is to review the assessment presented. In the course of this review, the Chair may request a reasonable number of sensitivity runs, additional details of the existing assessments, or similar items from technical staff. However, the Review Panel is not authorized to conduct an alternative assessment or to request an alternative assessment from the technical staff present. The Review Panel should outline in its report any remedial measures that the Panel proposes to rectify shortcomings in the assessment.

### **Specific Activities and Responsibilities**

The CIE shall provide a Chair and three Review Panelists to conduct the review of the Chesapeake Bay blue crab stock assessment.

### **Tasks**

It is estimated that the Chair's duties will occupy a total of 17 days - several days prior to the Review Panel meeting for document review; 3 days at the review meeting in Annapolis; several days following the meeting to ensure that the final documents are completed; and several days following the meeting to review the individual panelist's Review Reports and produce the Summary Report. This report shall be a summary of the individual Panelist Review Reports, accurately and fairly representing all viewpoints. There shall be no attempt by the Chair to develop a consensus report.

Roles and responsibilities:

- (1) Prior to the meeting: review the Chesapeake Bay blue crab assessment report and other relevant documentation in support of this review.
- (2) During the meeting: act as chairperson, where duties include control of the meeting, coordination of presentations, control of document flow, and facilitation of discussion.
- (3) After the meeting: provide a Summary Report, which summarizes the findings of the individual panelist's Review Reports (to be provided to the chair no later than August 25, 2005). The Summary Report shall be organized like the Review Reports, with an executive summary, a review of activities and a summary of findings and recommendations that collectively emerged from the meeting. Advice on additional questions that are directly related to the assessment and are raised during the meeting should be included in the report text. See Annex 1 for further details on report contents. The Chair shall not attempt to reach or describe consensus on the assessment, but shall fairly summarize the individual Review Reports and draw attention to the collective conclusions and recommendations.

The milestones and schedule for the Chair are summarized in the table below. The Chair shall begin the summarization using the draft individual Review Reports provided by the Panelists on August 25, 2005. When these individual panelist reports are finalized, following the CIE internal review and approval by the NMFS Contracting Officer's Technical Representative (COTR), the CIE

shall provide copies of the final versions to the Chair on September 12, 2005 for completion of the Summary Report. No later than September 20, 2005, the Chair shall submit the Summary Report<sup>1</sup> to the CIE. This shall be addressed to the “University of Miami Independent System for Peer review,” and sent to Dr. David Sampson, via e-mail to [david.sampson@oregonstate.edu](mailto:david.sampson@oregonstate.edu), and to Manoj Shivlani, via e-mail to [mshivlani@rsmas.miami.edu](mailto:mshivlani@rsmas.miami.edu). The CIE shall provide the final Summary Report to the NMFS COTR for final approval on September 27, 2005.

**Milestones or Report Delivery Dates**

The following table provides the milestones and delivery dates for conducting the panel review of the Chesapeake Bay blue crab stock assessment.

<b>Milestone</b>	<b>Date</b>
Panel review meeting in Annapolis, MD	August 9-11, 2005
Individual panelists provide their draft reports to CIE for review and to Chair for initiating development of the Summary Report	August 25, 2005
CIE provides reviewed individual panelist reports to NMFS COTR for approval	September 1, 2005
COTR notifies CIE of approval of individual panelist reports	September 8, 2005
CIE provides final individual panelist reports to COTR (with signed cover letter) and to Chair to complete Summary Report	September 13, 2005
Chair provides CIE with draft Summary Report for review	September 20, 2005
CIE provides reviewed Summary Report to COTR for approval	September 27, 2005
COTR notifies CIE of approval of Summary Report	September 30, 2005
CIE provides final Summary Report with signed cover letter to COTR	October 5, 2005
COTR provides final Summary Report to NEFSC contact	October 7, 2005

No consensus opinion among the CIE reviewers is sought, and all reports will be the product of the individual CIE reviewer or chairperson.

*NOAA Contact person:*

Derek Orner, NOAA Chesapeake Bay Office, 410 Severn Avenue, Annapolis, MD 21403; [Derek.ornier@noaa.gov](mailto:Derek.ornier@noaa.gov)

## **Appendix 2: Description of Chair's Summary Report**

### **ANNEX 1: Contents of Chair Summary Report**

1. The summary report shall summarize the findings of the individual panelist's Review Reports. The Chair shall not attempt to reach or describe consensus on an assessment, but shall fairly summarize the individual Review Reports and draw attention to the collective conclusions and recommendations.
2. The summary report shall be prefaced with an executive summary of findings and/or recommendations.
3. The main body of the report shall consist of a review of activities and, for each assessment reviewed, a summary of findings and recommendations that collectively emerged from the meeting.

### Appendix 3: Assessment Model Analysis (developed by P. Medley)

#### Objective of the Analysis

The objective of this analysis was to review the modelling approach adopted by the assessment and in particular the coherence among the various indices available to monitor the population size. This is not a new assessment, but rather allows more detailed exploration of the issues regarding the use of the various survey indices. Three models were fitted for comparison. The same modified Collie-Sissenwine population model was used as in the assessment, although the fitting procedure is slightly different. An extension of this model with a density dependent term modelling the effect of adult cannibalistic mortality on recruits was also applied. Finally, a simpler recruitment index model was fitted for comparison. In all cases I used an alternative and simpler minimising routine (Solver in MS Excel), assuming process error was zero. This latter change simplifies the fitting process while still allowing the indices to be assessed.

#### Fitted Models

The recruitment index model is very similar to the Collie-Sissenwine model, except a recruitment index is assumed to be exactly proportional to the recruitment in each season.

$$N_{t+1} = (N_t + \lambda r_t - C_t) e^{-M}$$

where  $N_t$  = population at time  $t$ ,  $\lambda$  = parameter raising the index to the population size,  $M$  = natural mortality and  $C_t$  = catches taken in time period  $t$ . The model assumes the recruitment index has no observation error. The longest time series index for age 0 animals, the VIMS trawl index, was found to give the best results and was used in the final model.

The CMS model allows each recruitment to be fitted as a separate value.

$$N_{t+1} = (N_t + R_t - C_t) e^{-M}$$

The  $R_t$  were fitted parameters to the model, as well as the  $N_0$ . Fitting the recruitment time series did not require recruitment to be exactly proportion to the recruitment indices, which allows for observation error. Because the model is deterministic (no process error), the remaining population time series values ( $N_t$ ) would be fixed by these parameters.

Finally, as an exploratory analysis, a population model was fitted which attempted, somewhat crudely, to capture density-dependent cannibalistic effects of adults on recruits. In this case, survival of recruits to the next year depends upon the number of adults.

$$N_{t+1} = (N_t + R_t e^{-\alpha N_t} - C_t) e^{-M}$$

The parameters in each model were fitted to the indexes by minimising the squared difference between the expected and observed log-indices.

$$\hat{I}_{it} = Ln(q_i N_t)$$

$$\text{Minimise } \sum (I_{it} - \hat{I}_{it})^2$$

Eight indices as well as the catch time series were used. All indices were transformed to natural logarithms of the original data and weighted equally. The appropriateness of the log-transform was not explored.

The  $q$  coefficients could be estimated directly through linear regression, so only the  $R_t$  and  $N_0$  parameters were estimated using the non-linear minimizer. This is broadly the same approach as used in the CMS model presented in the assessment.

**Results and Discussion**

The model failed to find a reasonable least squares solution unless the scaling of the ( $q_i$ ) parameters was fixed. The winter dredge survey biomass estimates were used to estimate the biomass directly (fixed  $q$  parameter for this index to 1.0). This was equivalent to using the exploitation fractions in the fitting process in the original assessment.

The general results follow the same pattern as those obtained from the fitted model in the 2005 assessment. However, in all cases, the estimated fishing mortality is lower than that obtained from the 2005 assessment. The model consistently estimated higher abundance than indicated by the WDS biomass estimate.

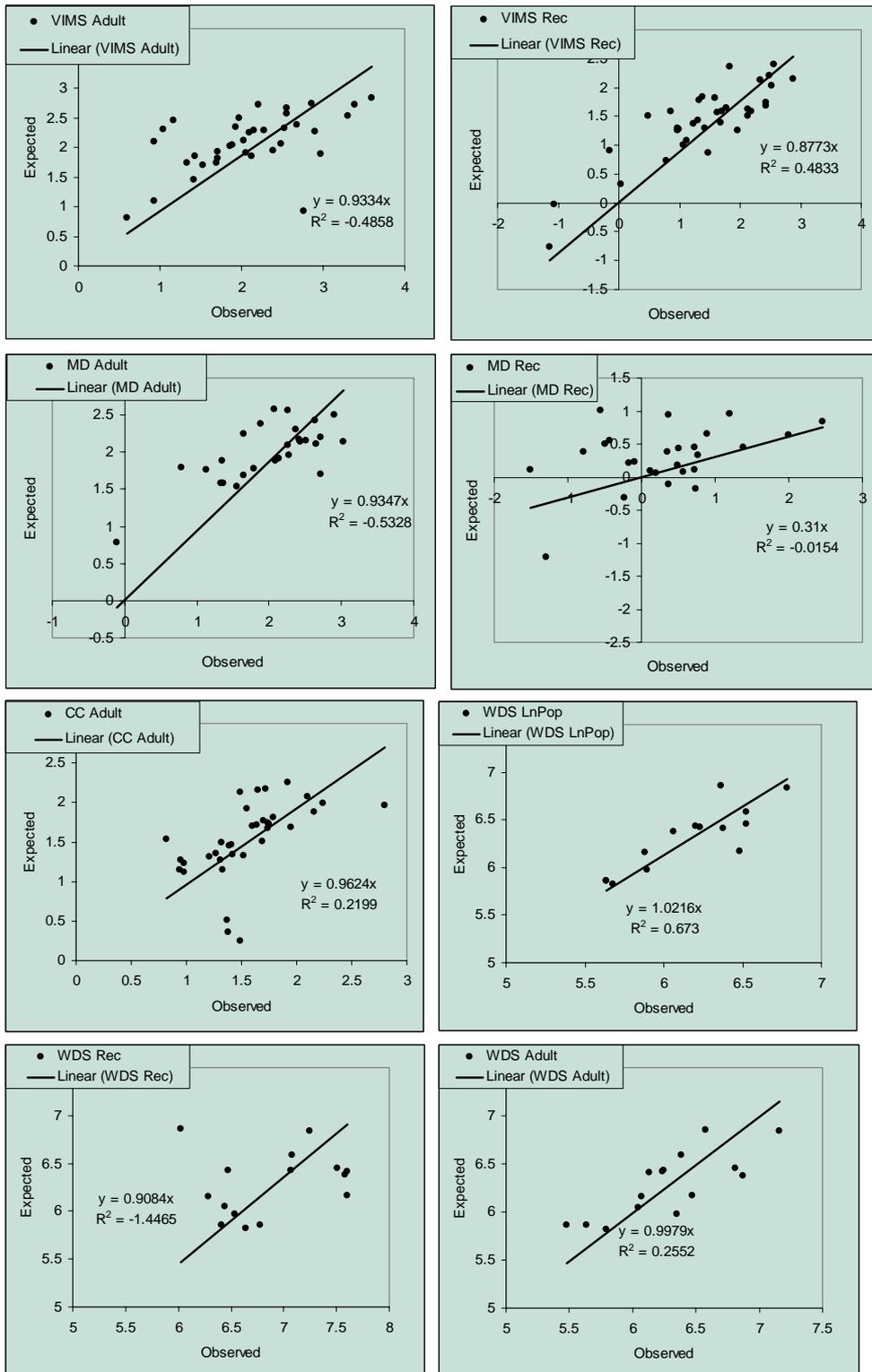
The better fit of the density dependent model indicated that there may be significant improvements in assessment models as the population processes are better modelled. The low fishing mortality in this case is partly an artifact of higher recruitment with its associated higher mortality. An appropriate selectivity model and within season mortality could lead to more realistic fishing mortality estimates.

In all models, the residuals were still significantly autocorrelated, suggesting that more of the time series behaviour could be captured by a population model. It was also found that the indices are not coherent and do not all possess the same signal (Fig. A3.1). As well as improving the standardization of the indices using GLMs, appropriate transforms and weights need to be critically examined to try to improve the use of these indices. This would include considering the timing and distribution of the stock relative to when and where the surveys are conducted. It may be that improvements in the indices will help choose between population models.

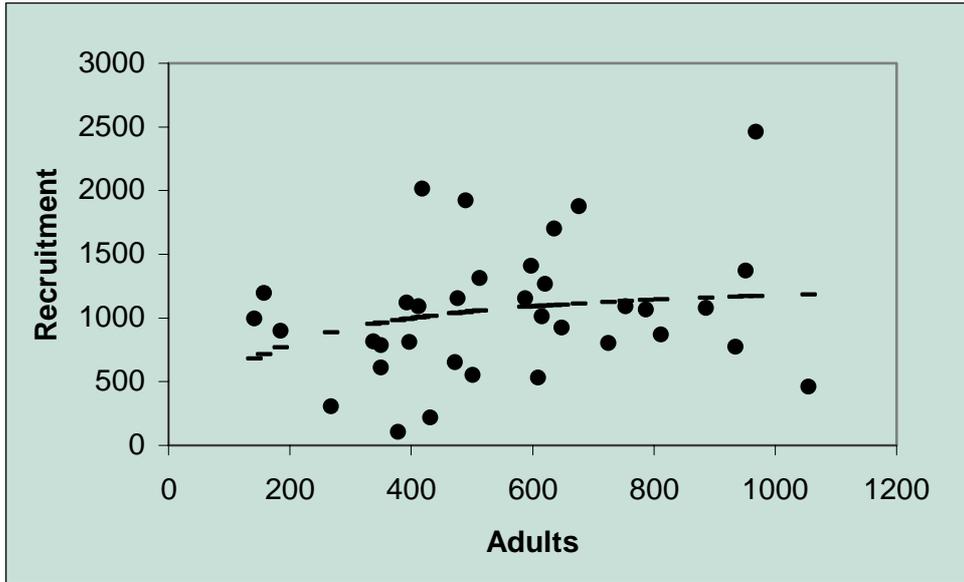
The results also show a hint of a workable stock-recruitment relationship (Figure A3.2). Improvements in the data and modelling would hopefully produce a more reliable functional relationship. As is usual with these relationships, the observation and process error predominates. However, such a relationship can be used as the default hypothesis on which to build a workable control rule that the managers and fishermen should find acceptable.

**Table A3.1** Summary of results from fitting the three models to the catches and available indices. The degrees of freedom (number of data points – number of parameters) remains the almost same for each model. Natural mortality was estimated for the density dependent Collie-Sissenwine model, but otherwise fixed at 0.9 year<sup>-1</sup>. The density dependent Collie-Sissenwine model fitted the indices best, but exhibits the lowest impact from the fishery with the lowest F.

Model	Degrees of Freedom	Goodness of Fit ( $\chi^2$ )	Natural Mortality (M)	Fishing Mortality (F) (geometric mean)
Recruitment Index	160	12.27	0.9	0.16
Collie-Sissenwine (No process error)	160	6.88	0.9	0.27
Collie-Sissenwine with density dependent mortality	159	5.85	0.7	0.06



**Figure A3.1** Observed-expected plots for VIMS and MD trawl indices the Collie-Sissenwine model with observation error only fitted in MS Excel. Where positive, the  $R^2$  statistic indicates the proportion of variance explained by the model. A slope close to 1.0 indicates low bias and general good fit. In general, the VIMS trawl, Winter dredge and Calvert Cliffs (CC) indices seemed more coherent, so that variation could be explained by simple changes in population size in these indices. The Maryland trawl index (MD) generally seemed to be the worst behaved agreeing with comments made in the 2005 assessment report.



**Figure A3.2** Approximate Beverton and Holt stock recruitment relationship fitted to the recruits and spawning stock numbers estimated from the modified Collie-Sissenwine model.

## Appendix 4. GLM for Depletion Experiments (developed by P. Medley)

Using an analysis of covariance approach, the x-variables are the cumulative catches within each catch category.

$$U_{it} = a_{it} - q_i \sum_{s=0}^{t-1} C_{is}$$

Both  $a_{it}$  and  $q_i$  must be fitted within each  $i$  category. The intercept and the slope are related as  $q$  occurs in the intercept. Where the  $q$  is not varied based on common covariates, the intercept will still need to change between experiments. Categories would be defined based on shared covariates, such as weather conditions, benthic habitat vessel and gear. The advantage of this approach is that the error term is shared and the significance of parameters can be tested leading to greater parsimony.

Keeping the model to a strict linear format prevents improved modelling. It is possible to restate the model in a more flexible near-linear form. The model can be defined as:

$$N_{it} = N_{i0} - \sum_{s=0}^{t-1} C_{is}$$

which gives the population size within each experiment. The catchability and selectivity can then be modelled with any number of covariates using a standard GLM. For example, a log-linear model could have the form:

$$\text{Ln}(\mu_i) = \text{Ln}(N_{it}) + lp_i$$

where  $lp_i$  is the appropriate linear predictor for category  $i$  and  $\mu_i$  is the expected CPUE or catch for the survey point as appropriate. The fitting follows an iterative two phase process. Iteratively, fitting the  $N_{i0}$  in the population model can be done first, then the GLM can be fitted in the normal way. Because of the near linearity of the model, fitting is straightforward using standard regression techniques. The same technique can also be applied with several populations, split in size groups for example, to estimate simultaneously catchability and selectivity curves.

The error model should be critically evaluated. It is quite possible that any of a normal, Poisson, binomial, gamma or log-normal may provide the best results and all should be evaluated particularly with respect to the variance-mean relationship. The log-normal and gamma models should generally be avoided if possible as they are undefined where catches are zero and delta-models, where a separate binomial likelihood is estimated for zero catches, are more complex. Zero catches in catch categories become a distinct possibility where the categories are broken down. Using a Poisson or binomial error under quasi-likelihood assumptions would be preferable.

Further improvements could be made to the model by considering more detailed spatial models of depletion, if GPS information is available. This would allow a direct estimate of dredge efficiency, that is the proportion of crabs that are captured which are in the path of the dredge. This allow a better direct estimate of density and thereby a better direct estimate of biomass.